For Reference

NOT TO BE TAKEN FROM THIS ROOM

For Reference

NOT TO BE TAKEN FROM THIS ROOM

Ex LIBRIS
UNIVERSITATIS
ALBERTAENSIS

UNIVERSITY OF ALBERTA
QUAECUMQUE VERA

University of Alberta
Printing Department

Digitized by the Internet Archive
in 2018 with funding from
University of Alberta Libraries

https://archive.org/details/Mailloux1963

THE UNIVERSITY OF ALBERTA

NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS

by

Barry J. Mailloux

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE.

DEPARTMENT OF MATHEMATICS

EDMONTON, ALBERTA

SEPTEMBER, 1963

ABSTRACT

This thesis is concerned with numerical approximations to the solutions of differential equations.

The elements of the finite-difference calculus are presented, and employed to develop several useful formulas. Elementary methods for the solution of differential equations are presented first, followed by more sophisticated techniques. Instability of difference approximations to initial value problems is discussed at some length, and an algorithm which attempts to minimize this difficulty is presented.

The classification of partial differential equations into hyperbolic, parabolic, and elliptic types is shown, and methods for the solution of all three types are given.

A discussion is given, approaching the Green's function from a numerical point of view. It is shown that this perspective leads to new insight into the subject. An analytic expression for the Green's function of a large class of differential operators is derived by an essentially numerical technique.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor R.S. Julius for his help and guidance in the writing of this thesis.

It is my great pleasure also to thank Dr. J. McNamee, who has acted as my faithful mentor during the past four years.

I thank Mrs. V.J. Manly who performed the task of typing this thesis.

Finally, I wish to acknowledge my debt to the Computing Center, and its staff for its help with this thesis, and for having provided a stimulating atmosphere in which to work.

TABLE OF CONTENTS

		Page
CHAPTER I	INTRODUCTION	1
Section 1.1	A Survey of the Thesis	1
Section 1.2	Notation Used in the Thesis	4
CHAPTER II	THE CALCULUS OF FINITE DIFFERENCES	6
Section 2.1	Introduction	6
Section 2.2	Finite Difference Operators	8
Section 2.3	Derivatives in Terms of Difference Operators	13
Section 2.4	Comparison of Finite Difference with Differential Calculus	16
Section 2.5	Finite Difference Equations	23
CHAPTER III	ELEMENTARY METHODS FOR THE NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS	29
Section 3.1	Introduction	29
Section 3.2	The Taylor Series Method for Initial Value Problems	31
Section 3.3	Euler's Method for Initial Value Problems	38
Section 3.4	Cauchy's Existence Theorem for Initial Value Problems	42
Section 3.6	Simple Approximations for Boundary Value Problems	53
Section 3.7	Classification of Second-Order Partial Differential Equations	60
Section 3.8	The Method of Characteristics for Hyperbolic Systems	65

		Page
Section 3.9	Solution of Parabolic Systems	72
Section 3.10	Solution of Elliptic Systems	74
CHAPTER IV	ADVANCED NUMERICAL METHODS FOR ORDINARY DIFFERENTIAL EQUATIONS	77
Section 4.1	Introduction to Multi-Step Formulas	77
Section 4.2	Milne's Method	85
Section 4.3	Stability of Multi-Step Methods	92
Section 4.4	Runge-Kutta Methods for Initial Value Problems	100
Section 4.5	Further Techniques for Boundary Value Problems	110
CHAPTER V	A NUMERICAL APPROACH TO THE GREEN'S FUNCTION	114
Section 5.1	Introduction	114
Section 5.2	A Proof of Convergence	121
Section 5.3	An Analytic Expression for the Green's Function	127
Section 5.4	Special Cases of the General Formula	136
CHAPTER VI	CONCLUDING REMARKS	142
BIBLIOGRAPHY		144

LIST OF TABLES AND FIGURES

		Page
TABLE I	Solution of $u(x+2h) = 2u(x+h)-(1-h^2)u(x)$	49
FIGURE I	The Characteristic Curves of a Hyperbolic Partial Differential Equation	67

CHAPTER I

Introduction

Section 1.1 A survey of the Thesis

One of the important problems common to all of the physical sciences is the formulation and solution of differential equations. A large, and still rapidly growing, literature has been developed in reponse to this need.

As might be expected, the passage of time has seen many of the less difficult problems solved, and so, the problems now posed are often quite difficult. More and more often, new problems are intractable by traditional techniques. For instance, it is known that no expression in terms of elementary functions can be given for the solutions of the equations

$$y' = x^2 + y^2, \qquad (1.1.1)$$

and

$$y'' = 6y^2 + x.$$

Furthermore, even if an expression is available for the solution of a differential equation, its evaluation may prove quite tedious; an example of this situation is given in Chapter III.

For many years, numerical methods for solving

differential systems in an approximate sense have been known, but, on account of the large amounts of arithmetic involved in carrying them out, they have not been extensively employed until recently.

Since the early and middle 1940's, electronic digital computers have become available. Currently available machinery can perform upwards of 250,000 arithmetic operations per second, and at a relatively modest cost (something like .01¢ per operation for an IBM 7090). Thus, methods of numerical approximation have been rendered practical, and therefore a great deal of recent research has been directed toward the further development of these techniques.

This thesis attempts to review the major techniques developed for the numerical integration of differential, and especially ordinary differential, equations. It does not pretend to be complete, since, to do so, it would have to be many times its present size; neither does it go into the detailed development of a topic, unless this is in some way revealing of general techniques.

The second chapter presents the basic results of the finite-difference calculus, and compares it with the more familiar differential calculus. Several results are derived for later use in the thesis.

Chapter III attempts, by way of elementary techniques and examples to provide a taste of the numerical approximation of differential equations. The purpose is to exhibit ideas, free from too much obscuring detail. Elementary approximations to initial and boundary value ordinary differential equations, and to hyperbolic, parabolic, and elliptic partial differential equations are presented and discussed. The problem of stability is illustrated by means of numerical example.

In Chapter IV, practical, and more advanced methods are discussed, for ordinary differential systems only. The important multi-step integration techniques are developed, and their stability properties are discussed in some detail. Runge-Kutta methods are also discussed, and programs for multi-step and Runge-Kutta integrations are given. Methods for two-point boundary-value problems are mentioned briefly.

In Chapter V, we pursue a line of thought partially developed in Chapters II and III. We consider the matrix equation obtained by replacing a differential equation by a finite-difference approximation. An expression is obtained for the solution of the algebraic problem, and it is shown that, in the limit as the tabular interval approaches zero,

the finite-difference approximation yields an exact solution in terms of Green's functions. Further an expression for the Green's function of a quite general differential operator is obtained by this method. It is believed that some of the results of this chapter are new.

The final chapter summarizes the major results of the thesis, and indicates some topics on which current research is proceeding.

Each chapter is preceded by a brief synopsis.

Section 1.2 Notation Used in the Thesis

The notation used is largely standard, the major deviation occurring in the programs of Chapters III and IV. These programs have been written using a subset of an algorithmic language proposed by Iverson [27].

According to this notation, ordinary variables are written as lower-case; vectors as underscored, lower-case; and matrices as underscored, upper-case Roman characters. An element of a vector is indicated by a subscript, without the underscore; a column vector of a matrix is indicated by a subscript, and the underscore is retained.

Each step of a program is numbered, for possible further reference. The notation

$$\alpha \leftarrow \beta,$$

implies that the variable α receives the value of the expression β. The form

$$\alpha:\beta,(r_1,r_2,\ldots)\rightarrow(s_1,s_2,\ldots)$$

means that α and β are compared, and that if α stands in the relation r_1 to β, control passes to step s_1; if in relation r_2, to s_2, etc. The relations are of the form $\leq$, $>$, $\neq$, and so on.

CHAPTER II

THE CALCULUS OF FINITE DIFFERENCES

This chapter introduces the finite difference calculus. It attempts to put it in a perspective which makes it seem as natural a descriptive device as differential calculus.

The elementary finite difference shifting, differencing, and averaging operators are defined, and some formal identities are obtained.

Expansions of the familiar differentiation operator in terms of differencing operators are derived.

Some formal analogies between finite difference and differential calculi are shown. In particular, an example is given of the solution of a difference equation by techniques borrowed directly from differential calculus.

Section 2.1 Introduction

In the elementary study of differential calculus, we may approach the concept of a derivative in the following way:

We have an intuitive notion of what is meant by the "slope" of a given curve $f(x)$ at any given point x_o. We can easily speak of the "average slope," $S(x)$ of $f(x)$ in the interval, say, $x_o \leq x \leq x_1$, defined as

$$S\left(\frac{x_o+x_1}{2}\right) = \frac{f(x_1) - f(x_o)}{x_1 - x_o} . \qquad (2.1.1)$$

By, "the slope of $f(x)$ at x_o" we then understand

the limit (it it exists) of (2.1.1) as x_1 is made to approach x_o. This limit is termed the derivative of $f(x)$ at $x = x_o$. The derivative of $f(x)$ at any x is written

$$f'(x) = \frac{df(x)}{dx} = Df(x) = \lim_{x_1 \to x} \frac{f(x_1) - f(x)}{x_1 - x} . \qquad (2.1.2)$$

It is worth noting, in passing, that this definition requires at least the continuity of $f(x)$ at x, and that, in the physical world, this condition is seldom fulfilled, due to the atomicity of most material objects. Thus, in the real world, this concept is usually an idealisation, since the limit of (2.1.1) cannot actually be taken.

As an example of this, the derivation of the well known string equation proceeds by considering the effect of elastic forces on a very short section of string, applying Newton's second law of motion, and deriving an equation. The length of the section of string is then reduced without limit, and a differential equation results. Really, this process is not valid, since strings are known to be composed of atoms and molecules of non-zero size. However, for practical purposes, i.e., for the sake of a manageable physical model, this distinction is trivial.

Nevertheless, we have illustrated here the important fact that most of the differential equations used in the physical sciences are derived by obtaining an equation at separated points and then letting the points approach one another. This thesis will explore some of the implications of this procedure, especially in Chapter V.

Again in passing, it is to be admitted that some differential equations are not obtained in this way. For instance, in quantum mechanics, Schrödinger's wave equation is obtained essentially by inspection, experiment, and inspiration. This is also the case for Charles' law for gases and Newton's laws of motion.

Let us now go on to the finite-difference calculus itself.

Section 2.2 Finite Difference Operators

We consider a function, say, $f(x)$, defined at the abscissae

$$x = x_0, x_1, x_2, \ldots, x_n,$$

where, for some real number h (usually, but not necessarily, thought of as being small),

$$x_i = x_0 + ih \quad ; \quad i = 0,1,2,\ldots,n.$$

We will usually write

$$f(x_i) = f_i.$$

A number of operations can now be defined on the f_i. The names of the operations, the symbols used to represent them, and the effect of each operator, acting on f, follow:

Forward Shift operator, E , $Ef_i = f_{i+1}$ (2.2.1)

Forward Difference operator, Δ , $\Delta f_i = f_{i+1} - f_i$ (2.2.2)

Backward Difference operator, ∇ , $\nabla f_i = f_i - f_{i-1}$ (2.2.3)

Central Difference operator, δ , $\delta f_i = f_{i+1/2} - f_{i-1/2}$ (2.2.4)

Averaging operator, μ , $\mu f_i = 1/2(f_{i+1/2} + f_{i-1/2})$ (2.2.5)

Differentiation operator, D , $Df_i = \left(\frac{df(x)}{dx}\right)_{x=x_i}$ (2.2.6)

Superscripts are used to indicate repeated application of an operator. That is, representing any of the above operators by O,

$$O^n f_i = O(O^{n-1} f_i). \tag{2.2.7}$$

For example,

$$E^2 f_i = E(Ef_i) = E(f_{i+1}) = f_{i+2}, \tag{2.2.8}$$

and

$$E^n f_i = f_{i+n}. \tag{2.2.9}$$

We will treat these operators formally, to obtain some useful identities. The rigorous justification for most of these identities can only be given a posteriori, and in the context of a particular function, f(x), or a particular class of functions.

First, we will express the shift operator, E, in terms of the differentiation operator, D. We have

$$Ef_i = f_{i+1};$$

that is,

$$Ef(x_i) = f(x_{i+h}). \tag{2.2.10}$$

But, by Taylor's theorem,

$$f(x_i+h) = f(x_i) + hDf(x_i) + \frac{h^2D^2f(x_i)}{2!} + \ldots \quad . \tag{2.2.11}$$

That is,

$$f(x_i+h) = (1+hD+\frac{h^2D^2}{2!} + \frac{h^3D^3}{3!} + \ldots)f(x_i). \tag{2.2.12}$$

We recognize the series in (2.2.12) as the formal series for the exponential function. Hence, we write

$$Ef_i = f(x_i+h) = e^{hD}f(x_i) = e^{hD}f_i, \tag{2.2.13}$$

and thus, we obtain the formal relation

$$E = e^{hD}. \tag{2.2.14}$$

Now, we will express the three difference operators and the averaging operator in terms of E.

We have

$$\Delta f_i = f_{i+1} - f_i = Ef_i - f_i = (E-1)f_i, \tag{2.2.15}$$

and hence, formally,

$$\Delta = E-1. \tag{2.2.16}$$

Now,

$$\nabla f_i = f_i - f_{i-1} = f_i - E^{-1}f_i = (1-E^{-1})f_i, \tag{2.2.17}$$

and, therefore,

$$\nabla = (1-E^{-1}). \tag{2.2.18}$$

We note that E^{-1} is a "backward shift operator."

Also,

$$\delta f_i = f_{i+1/2} - f_{i-1/2} = E^{1/2}f_i - E^{-1/2}f_i$$

$$= (E^{1/2}-E^{-1/2})f_i, \tag{2.2.19}$$

and thus,

$$\delta = (E^{1/2}-E^{-1/2}). \tag{2.2.20}$$

Finally,

$$\mu f_i = 1/2(f_{i+1/2}+f_{i-1/2}) = \frac{1}{2}(E^{1/2}+E^{-1/2})f_i, \qquad (2.2.21)$$

and so

$$\mu = \frac{1}{2}(E^{1/2}+E^{-1/2}). \qquad (2.2.22)$$

We can combine these relations with (2.2.14) to obtain expressions in terms of D.

We have:

$$\Delta = E-1 = e^{hD}-1; \qquad (2.2.23)$$

$$\nabla = 1-E^{-1} = 1 -e^{-hD}; \qquad (2.2.24)$$

$$\delta = (E^{1/2}+E^{-1/2}) = (e^{hD/2}-e^{-hD/2})=2 \sinh \frac{hD}{2}; \qquad (2.2.25)$$

and

$$\mu = \frac{1}{2}(E^{1/2}+E^{-1/2}) = \frac{1}{2}(e^{hD/2}+e^{-hD/2}) = \cosh \frac{hD}{2} . \qquad (2.2.26)$$

A further relation which we will need later can now be derived. Substituting (2.2.25) and (2.2.26) into the identity

$$\cosh^2 \tfrac{1}{2}hD - \sinh^2 \tfrac{1}{2}hD = 1, \qquad (2.2.27)$$

we obtain,

$$\mu^2 - \frac{1}{4}\delta^2 = 1. \qquad (2.2.28)$$

i.e.,

$$\mu = (1 + \tfrac{1}{4}\delta^2)^{1/2}. \qquad (2.2.29)$$

In fact, each of the operators, (2.2.1) to (2.2.6), can be expressed in terms of each of the others [1], but the relations derived above are representative, and will suffice for our purposes.

Section 2.3 Derivatives in Terms of Difference Operators

From the point of view of solving differential equations, it is very useful to be able to express the operator D, and its powers, as evaluable expressions of difference operators. Our motivation in this section is partly formal and theoretical, and partly practical.

Writing (2.2.23) as

$$e^{hD} = 1 + \Delta, \tag{2.3.1}$$

and taking natural logarithms of both sides, we obtain

$$hD = \ln(1+\Delta). \tag{2.3.2}$$

This is, itself, not an evaluable expression in terms of Δ, but we can make it so by expressing the right-hand side as a Taylor series. Then,

$$hD = \ln(1+\Delta) = (\Delta - \tfrac{1}{2}\Delta^2 + \tfrac{1}{3}\Delta^3 + \ldots). \tag{2.3.3}$$

In fact, for derivatives of any order, m, we have

$$h^m D^m = (\Delta - \tfrac{1}{2}\Delta^2 + \tfrac{1}{3}\Delta^3 + \ldots)^m. \qquad (2.3.4)$$

In particular, for m = 2, [2],

$$h^2 D^2 = (\Delta^2 - \Delta^3 + \tfrac{11}{12}\Delta^4 - \tfrac{5}{6}\Delta^5 + \ldots). \qquad (2.3.5)$$

In a similar fashion,

$$h^m D^m = (\nabla + \tfrac{1}{2}\nabla^2 + \tfrac{1}{3}\nabla^3 + \ldots)^m \qquad (2.3.6)$$

Now, rewrite (2.2.25) as,

$$hD = 2 \sinh^{-1} \delta/2. \qquad (2.3.7)$$

i.e.,

$$\left(\frac{hD}{\delta}\right)^m = \left(\frac{\sinh^{-1} \delta/2}{\delta/2}\right)^m. \qquad (2.3.8)$$

The Taylor series for $\sinh^{-1}x$ is, [3],

$$\sinh^{-1}x = x - \tfrac{1}{6}x^3 + \tfrac{3}{40}x^5 - \tfrac{5}{112}x^7 + \ldots \quad . \qquad (2.3.9)$$

Hence,

$$h^m D^m = \delta^m \left(1 - \frac{\delta^2}{24} + \frac{3}{640}\delta^4 - \frac{5}{14{,}336}\delta^6 \ldots\right)^m . \qquad (2.3.10)$$

In particular, for m = 2,

$$h^2D^2 = \delta^2\left(1 - \frac{\delta^2}{12} + \frac{\delta^4}{90} - \frac{\delta^6}{560} + \ldots\right). \qquad (2.3.11)$$

We notice that the coefficients in (2.3.10) become small more rapidly than those of (2.3.4) or (2.3.6). That is, (2.3.10) tends to converge more rapidly than the others. This seems to be a typical feature of central difference formulae in general.

One objectionable feature of (2.3.10) is that for odd values of m, the right-hand side contains only odd powers of δ. Hence, in order to use this formula, we would need to know our function f(x) at the points $x_o + \frac{1}{2}h$, $x_o + \frac{3}{2}h$, and so on. To see this, we write (2.2.20):

$$\delta = E^{1/2} - E^{-1/2} = E^{-1/2}(E-1). \qquad (2.3.12)$$

Therefore,

$$\delta^n = E^{-n/2}(E - 1)^n.$$

Clearly, for n odd, this series contains a half power of E in every term, and implies the need for values of f at the mid-tabular points.

To circumvent this difficulty, we multiply the

right-hand side of (2.3.10), by μ^m/μ^m, to obtain

$$h^m D^m = \mu^m \delta^m \left[\mu^{-1}\left(1 - \frac{\delta^2}{24} + \frac{3}{640}\delta^4 - \ldots\right)\right]^m . \qquad (2.3.13)$$

According to (2.2.29),

$$\mu^{-1} = \left(1 + \frac{1}{4}\delta^2\right)^{-1/2}$$

$$= 1 - \frac{\delta^2}{8} + \frac{3}{120}\delta^4 - \ldots \quad . \qquad (2.3.14)$$

Thus, multiplying the two series together, we obtain,

$$h^m D^m = (\mu\delta)^m \left(1 - \frac{1}{6}\delta^2 + \frac{1}{30}\delta^4 - \ldots\right)^m . \qquad (2.3.15$$

This series involves only the tabular points of $f(x)$, for even or odd m. It could be used in place of (2.3.10), even for even m, except for its somewhat inferior convergence.

Relations (2.3.10) and (2.3.15), as well as other generally useful derivative formulae, are tabulated in [2], for various values of m.

Section 2.4 Comparisons of Finite Difference with Differential Calculus

We can again consider the "average slope"

formula (2.1.1):

$$S\left(\frac{x_o+x_1}{2}\right) = \frac{f(x_1) - f(x_o)}{x_1-x_o} .$$

More generally,

$$S\left(\frac{x_i+x_{i+1}}{2}\right) = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}$$

$$= \frac{\Delta f_i}{\Delta x_i} . \qquad (2.4.1)$$

Thus,

$$f'(x)_{x=x_i} = Df_i = \lim_{\Delta x_i \to 0} \frac{\Delta f_i}{\Delta x_i} , \qquad (2.4.2)$$

and so, for small $\Delta x_i = h$, we expect that

$$Df_i \approx \frac{\Delta f_i}{h} ,$$

or, in other words, that

$$hD \approx \Delta . \qquad (2.4.3)$$

We are thus led to suspect that the differentiation operator, D, and the various difference operators, might, formally at least, behave similarly. In fact, they do so as we shall show, although there are differences which

prevent the analogy from being complete.

For instance, for suitable functions f and g, and constants, A and B,

$$D(Af + Bg) = ADf + BDg,$$

and

$$\Delta(Af_i + Bg_i) = A\Delta f_i + B\Delta g_i. \qquad (2.4.4)$$

In this case, the analogy is complete. In the following, we consider derivatives and differences of quotients and products, and the analogy is somewhat damaged.

In differential calculus,

$$D(f\cdot g) = f\cdot(Dg) + g\cdot(Df). \qquad (2.4.5)$$

In difference calculus,

$$\Delta(f\cdot g) = (E-1)(f\cdot g) = E(f\cdot g) - f\cdot g. \qquad (2.4.6)$$

Now, by definition,

$$E(f\cdot g) = f(x+h)g(x+h) = (Ef)\cdot(Eg). \qquad (2.4.7)$$

Hence,

$$\Delta(f\cdot g) = (Ef)\cdot(Eg) - f\cdot g$$

$$= (Ef)\cdot(Eg) + (Ef)\cdot g - (Ef)\cdot g + f\cdot g$$

$$= (Ef)\cdot(Eg-g) + (Ef-f)\cdot g$$

$$= (Ef)\cdot(\Delta g) + g\cdot(\Delta f). \qquad (2.4.8)$$

By a similar argument,

$$\Delta(f\cdot g) = f\cdot(\Delta g) + (Eg)\cdot(\Delta f) . \qquad (2.4.9)$$

Taking the average of (2.4.8) and (2.4.9), we can obtain the symmetric form:

$$\Delta(f\cdot g) = \frac{(E+1)}{2} f\cdot(\Delta g) + \frac{(E+1)}{2} g\cdot(\Delta f) \qquad (2.4.10)$$

The corresponding expressions for the backward difference operator are:

$$\nabla(f\cdot g) = (E^{-1}f)\cdot(\nabla g) + g\cdot(\nabla f) \qquad (2.4.11)$$

$$= f\cdot(\nabla g) + (E^{-1}g)\cdot(\nabla f) \qquad (2.4.12)$$

$$= \frac{(E^{-1}+1)}{2} f\cdot(\nabla g) + \frac{(E^{-1}+1)}{2} g\cdot(\nabla f). \qquad (2.4.13)$$

All of equations (2.4.8) to (2.4.13) are formally similar to (2.4.5) except for the operators involving E. We also note that for $h \to 0$, all of these operators involving E tend to unity.

Again in differential calculus,

$$D(f/g) = \frac{g\cdot(Df) - f\cdot(Dg)}{g^2} \qquad (2.4.14)$$

whereas, in difference calculus,

$$\Delta(f/g) = (E-1)(f/g) = E(f/g) - f/g$$

$$= (Ef)/(Eg) - f/g = \frac{g\cdot(Ef) - f\cdot(Eg)}{g\cdot(Eg)}$$

$$= \frac{g\cdot(Ef) - g\cdot f + g\cdot f - f\cdot(Eg)}{g\cdot(Eg)}$$

$$= \frac{g\cdot(Ef-f) - (Eg-g)\cdot f}{g\cdot(Eg)}$$

$$= \frac{g\cdot(\Delta f) - f\cdot(\Delta g)}{g\cdot(Eg)} \quad . \qquad (2.4.15)$$

Similarly,

$$\nabla(f/g) = \frac{g\cdot(\nabla f) - f\cdot(\nabla g)}{g\cdot E^{-1}g} \quad . \qquad (2.4.16)$$

Once again, (2.4.14) is formally the same as (2.4.15) and (2.4.16) except for the operators E and E^{-1}.

In differential calculus, the operation of integration is inverse to D. In difference calculus, summation is inverse to differencing, as we shall now show.

Consider the definite summation

$$\sum_{a}^{b} g(x) = g(a) + g(a+h) + g(a+2h)+\ldots+g(b), \qquad (2.4.17)$$

where g is a function, and a and b are real numbers such that $a<b$, and $(b-a) = nh$.

Suppose that we know some function such that

$$\Delta f(x) = f(x+h) - f(x) = g(x). \qquad (2.4.18)$$

In other words, f is an "anti-difference" of g, which we write

$$f(x) = \Delta^{-1} g(x). \qquad (2.4.19)$$

Substituting (2.4.18) into (2.4.17), we have,

$$\sum_{a}^{b}{}' g(x) = g(a)+ g(a+h) + g(a+2h)+\ldots+ g(b-h)+ g(b)$$

$$= \Big[f(a+h) - f(a)\Big] + \Big[f(a+2h) - f(a+h)\Big] + \Big[f(a+3h)-f(a+2h)\Big]$$

$$+\ldots+ \Big[f(b) - f(b-h)\Big]+\Big[f(b+h)-f(b)\Big]. \qquad (2.4.20)$$

We observe that all the terms on the right in (2.4.20) cancel, except for -f(a) and f(b+h). Hence,

$$\sum_{a}^{b}{}' g(x) = f(b+h) - f(a). \qquad (2.4.21)$$

This corresponds closely to the "fundamental theorem" of the integral calculus:

$$\int_a^b g(x)dx = f(b) - f(a), \text{ where } Df(x) = g(x) \qquad (2.4.22)$$

In fact, the method for obtaining (2.4.21) suggests a method of deriving (2.4.22). For, consider the definition of the definite integral,

$$\int_a^b g(x)dx = \lim_{h\to 0} \sum_{a}^{b}{}' hg(x) = \lim_{h\to 0} h\sum_{a}^{b}{}' g(x) \qquad (2.4.23)$$

Also,

$$Df(x) = \lim_{h \to 0} \frac{\Delta f(x)}{h} .$$

Furthermore, as $h \to 0$,

$$f(b+h) \to f(b).$$

The above argument does not, of course, even begin to be rigorous. It does, however provide some additional motivation for the "fundamental theorem" beyond that usually given in elementary calculus courses [6]; that is, in a sense, it shows "why" the theorem is true.

There are many more formal similarities between the difference and differential calculi [4] . We consider briefly difference and differential equations in the following section.

Section 2.5 Finite Difference Equations

In the differential calculus, the general form of an ordinary differential equation of order n is

$$F(D^n f, D^{n-1} f, \ldots, Df, f, x) = 0 \qquad (2.5.1)$$

where F is any function of n + 2 arguments, and f(x) is

an unknown function.

Corresponding to (2.5.1), the general difference equation of order n may be written

$$F(\Delta^n f, \Delta^{n-1} f, \ldots, \Delta f, f, x) = 0. \qquad (2.5.2)$$

Since we can always write difference operations in terms of shift operations, as

$$\Delta^k f = a_k E^k f + a_{k-1} E^{k-1} f + \ldots + a_1 Ef + a_o f, \qquad (2.5.3)$$

it is therefore clear that an alternate expression for (2.5.2) is

$$G(E^n f, E^{n-1} f, \ldots, Ef, f, x) = 0, \qquad (2.5.4)$$

or also,

$$G\left[f(x+nh),\ f(x+\overline{n-1}h), \ldots,\ f(x+h),\ f(x), x\right] = 0. \qquad (2.5.5)$$

In the sequel, purely for simplicity of notation, we will assume h = 1. It is clear that appropriate transformations can be made to difference equations with arbitrary h, to bring them to this form.

As we shall presently demonstrate, many of the elementary tricks and techniques for solving differential equations can be applied, with slight modifications, to difference equations.

For example, consider the difference equation

$$f(x+2) - 7\,f(x+1) + 10f(x) = 0. \qquad (2.5.5)$$

As a trial solution of (2.5.5.), we try $f(x) = a^x$. Substituting this trial function into the equation, we have

$$a^{x+2} - 7a^{x+1} + 10a^x = 0. \qquad (2.5.6)$$

i.e., $$a^x(a^2 - 7a + 10) = 0 \qquad (2.5.6)$$

Now, the case of $a = 0$ is a trivial solution, and we disregard it. The remainder of (2.5.6) is a quadratic in a, and has roots

$$a = 2 \text{ and } a = 5.$$

Hence, it seems that a general solution of (2.5.5) is

$$f(x) = A\cdot 2^x + B\cdot 5^x. \qquad (2.5.7)$$

That (2.5.7) is indeed a solution may be verified by direct substitution into (2.5.5).

However, it can also easily be verified that

$$f(x) = Au(x)2^x + Bv(x)5^x + w(x) \qquad (2.5.8)$$

is also a solution, for any $u(x)$ that has some constant value, for any $v(x)$ that has a constant value, and for any $w(x)$ that is zero, at all the tabular points.

For instance,

$$f(x) = A2^x\cos(2\pi x) + b5^x\tan(\pi x + \tfrac{\pi}{6}) + \sin\pi x \qquad (2.5.9)$$

is also a solution of (2.5.5), since $\cos(2\pi x)$ is 1, $\tan(\pi x+\frac{\pi}{6})$ is $\frac{1}{\sqrt{3}}$, and $\sin\pi x$ is zero at all the tabular points, $x = 0,1,2,\ldots$.

If, then, we are concerned only with the solution at tabular points, as is indeed the case in this thesis, then (2.5.7) provides an unique solution when A and B have been fixed.

Clearly, in order to determine the constants A and B, we require two additional conditions. This requirement is entirely analogous to the two conditions needed for a linear, second order, differential equation.

These conditions may take the form of one equation at each of two points (boundary conditions), or two independent equations at one point (initial conditions).

For example, we might have boundary conditions like

$$f(1) = 12, \text{ and } f(3) = 258. \qquad (2.5.10)$$

Hence, we could write the linear equations:

$$\begin{aligned} 12 &= 2A + 5B \\ 258 &= 8A + 125B, \end{aligned} \qquad (2.5.11)$$

which we can readily solve, to find,

$$A = 1, \; B = 2,$$

and hence,

$$f(x) = 2^x + 2\cdot(5^x). \qquad (2.5.12)$$

Alternatively, we might have had initial conditions, like,

$$f(1) = 9, \text{ and } \Delta f(1) = 24. \qquad (2.5.13)$$

Substituting the first of these into (2.5.7), we have

$$9 = 2A + 5B. \qquad (2.5.14)$$

Differencing (2.5.7), we have

$$\Delta f(x) = A\Delta(2^x) + B\,\Delta(5^x)$$

$$= A(2^{x+1}-2^x) + B(5^{x+1}-5^x)$$

$$= 2^x A + 4B5^x. \qquad (2.5.15)$$

Substituting the second of (2.5.13) into (2.5.15), we obtain

$$24 = 2A + 20B. \qquad (2.5.16)$$

From (2.5.14) and (2.5.16), we have, immediately,

$$A = 2, \text{ and } B = 1,$$

and hence,

$$f(x) = 2\cdot 2^x + 5^x = 2^{x+1} + 5^x. \qquad (2.5.17)$$

More difficult problems require additional results about antidifferences, and, in particular, about factorial polynomials and Stirling numbers [5]. However, the preceeding example will serve to illustrate the basic approach.

CHAPTER III

ELEMENTARY METHODS FOR THE NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS

This chapter employs the techniques of the previous chapter - and some others - for the solution of differential equations.

The Taylor series and Euler approximation methods for sets of first order initial value differential equations are developed. The important Cauchy existence and uniqueness theorem rising out of the Euler approximation is stated and discussed, and stability of initial value problems is illustrated by means of an example of an unstable approximation.

Boundary value ordinary differential equations, including eigenvalue problems, are briefly introduced.

The general, second-order, quasi-linear partial differential equation is classified into hyperbolic, parabolic, and elliptic types. The concept of characteristic curves is introduced and the solution of hyperbolic equations by the method of characteristics is described in detail. Simplified outlines are given of methods for solving parabolic and elliptic systems.

Section 3.1 Introduction

Increasingly, the new differential equations which arise in practice, do not admit of a solution in closed form. Even if they do, the labour involved in tabulating the solution may be quite large.

An example given in Modern Computing Methods [11] is illustrative:

The solution of the simple equation

$$\frac{dy}{dx} - \frac{2y}{1-x^4} = 0, \qquad (3.1.1)$$

can be shown to be

$$y = A\left(\frac{1+x}{1-x}\right)^{1/2} \exp(\tan^{-1}x). \qquad (3.1.2)$$

The very slightly more complicated equation

$$\frac{dy}{dx} - \frac{2y}{1-x^4} = x \qquad (3.1.3)$$

has the solution

$$y = \left(\frac{1+x}{1-x}\right)^{1/2} \exp(\tan^{-1}x)\left[\int x\left(\frac{1-x}{1+x}\right)^{1/2} \exp\right.$$

$$\left.(-\tan^{-1}x)\, dx + A\right]. \qquad (3.1.4)$$

To evaluate (3.1.4) is no mean task; in particular, exponential, arctangent, and square root functions must be evaluated, and a numerical quadrature performed.

The present chapter presents an alternative procedure - direct numerical approximation from the differential

equation. Our aim in this chapter is mostly to introduce principles; the methods described are not necessarily the methods to be used in everyday practice. More advanced procedures are discussed in the next chapter.

Section 3.2 The Taylor Series Method for Initial Value Problems

As a preliminary to this section, we first describe the reduction of higher-order ordinary differential equations subject to initial conditions, to sets of first order equations.

Although the most general form of an n'th order ordinary differential equation is that of (2.5.1), we will consider a slightly restricted form, namely

$$D^n y(x) = f(D^{n-1}y(x), D^{n-2}y(x), \ldots, Dy(x), y(x), x). \qquad (3.2.1.)$$

In other words, we assume that we can solve (2.5.1) formally for the highest derivative present.

Let us suppose that (3.2.1) is subject to the initial conditions

$$\begin{aligned} y(x_o) &= \eta_o \\ Dy(x_o) &= \eta_1 \\ &\vdots \quad \vdots \\ D^n y(x_o) &= \eta_n \end{aligned}$$

at the point $x = x_o$.

We write

$$\begin{aligned} y_o(x) &= y(x) \\ y_1(x) &= Dy(x) \\ y_2(x) &= D^2y(x) \\ &\vdots \qquad\qquad \vdots \end{aligned}$$

$$y_n(x) = D^n y(x) = f(y_{n-1}, y_{n-2}, \ldots, y_1, y_o, x) \qquad (3.2.3)$$

Then, differentiating the first of (3.2.3), we have

$$Dy_o = y_o' = y' = y_1.$$

That is

$$y_o' = y_1. \qquad (3.2.4)$$

We can similarly obtain

$$\begin{aligned} y'_1 &= y_2 \\ y'_2 &= y_3 \\ &\vdots \quad \vdots \end{aligned}$$

$$y'_{n-1} = y_n = f(y_{n-1}, y_{n-2}, \ldots, y_o, x). \qquad (3.2.5)$$

The initial conditions (3.2.2) can then be written

$$\begin{aligned} y_o(x_o) &= \eta_o \\ y_1(x_o) &= \eta_1 \\ &\vdots \qquad \vdots \\ y_n(x_o) &= \eta_n. \end{aligned} \qquad (3.2.6)$$

It is then clear that the system

$$\begin{aligned} y'_o(x) &= f_o(x,y_o,y_1,y_2,\ldots,y_n) \\ y'_1(x) &= f_1(x,y_o,y_1,y_2\ldots,y_n) \\ &\vdots \qquad\qquad \vdots \\ y'_n(x) &= f_n(x,y_o,y_1,y_2,\ldots y_n) \end{aligned} \tag{3.2.7}$$

subject to initial conditions (3.2.6), includes (3.2.5), and, therefore, we are, in effect, studying the solution of (3.2.1) and (3.2.2).

As an example of this reduction, consider the non-linear equation

$$g''(x) = g(x) - \frac{g^3(x)}{x^2}\ , \tag{3.2.8}$$

subject to

$$\begin{aligned} g(0) &= 0 \\ g'(0) &= \eta. \end{aligned} \tag{3.2.9}$$

Writing

$$\begin{aligned} y_o(x) &= g(x) \\ y_1(x) &= g'(x), \end{aligned} \tag{3.2.10}$$

we obtain the system

$$y'_o(x) = y_1(x)$$
$$y'_1(x) = y_o(x)\left[1 - y^2{}_o(x)/x^2\right] \qquad (3.2.11)$$

subject to

$$y_o(0) = 0$$
$$y_1(0) = \eta. \qquad (3.2.12)$$

Returning our attention now to (3.2.7), we consider a typical equation, say the i'th. We have

$$y'_i(x) = f_i(x,y_o,y_1,\ldots,y_n). \qquad (3.2.13)$$

Differentiating with respect to x,

$$y''_i(x) = \frac{\partial f_i}{\partial x} + \frac{\partial f_i}{\partial y_o} y'_o + \frac{\partial f_i}{\partial y_1} y'_1 + \ldots + \frac{\partial f_i}{\partial y_n} y'_n \ . \qquad (3.2.14)$$

Assuming sufficient differentiability of f_i with respect to all of its arguments, we can go on to obtain expressions for derivatives of $y_i(x)$ of all orders. To do this, we must proceed recursively. Using (3.2.14), we obtain $y''_i(x)$ at $x = x_o$, for all values of i. These values we substitute into the expression for $y'''_i(x)$, obtaining $y'''_i(x_o)$ for all i, and so forth, for as many terms as we

need, or can reasonably obtain.

We will write $y_i(x)$ as a Taylor series; in particular, for $x = x_o+h$, we have

$$y_i(x_o+h) = y_i(x_o) + hy'_i(x_o) + \frac{h^2}{2!} y''_i(x_o) +\dots \quad . \tag{3.2.15}$$

Since the numbers $y_i(x_o)$ are given as initial conditions, and the values of all required derivatives of $y_i(x_o)$ are calculable as described above, we can obtain values for all the $y_i(x_o+h)$.

Obviously, we do not intend an infinite summation of derivatives in (3.2.15). Where we stop depends on two things: the difficulty of evaluating higher derivatives; and, the necessity of obtaining sufficient convergence with the chosen value of h.

For instance, consider the function

$$y = a \sin x + be^x, \tag{3.2.16}$$

which satisfies

$$y^{iv} = y. \tag{3.2.17}$$

Suppose the initial conditions are such that b is quite small. Then clearly, the solution behaves much like a

sine function for small x, like an exponential for large x, and is more complex for intermediate x. For a given number of terms in the Taylor expansion, it is clear that a different h will be required to obtain satisfactory convergence in each region; or, conversely, that for a given fixed h, a different number of terms in the Taylor expansion will be required.

In practice, we often discover that higher-order derivatives are quite difficult to evaluate, and so choose h small enough to quarantee convergence to the desired number fo figures, using only those derivatives which are evaluable with reasonable effort.

Once all the $f_i(x)$ have been advanced to $x=x_o+h$, we can regard them as new "initial conditions," and advance to the next point, usually x_o+2h. We continue thus until values of $f_i(x)$ have been obtained for as many values of x as desired.

It is desirable to estimate very carefully the amount of calculation needed to advance the solution. Higher derivatives often tend to have many terms, and thus are costly to evaluate. It will often happen that a decrease in the size of h, even at the cost of more points to be calculated, will give accuracy comparable with

the inclusion of another derivative term, and at less total cost in effort of evaluation.

Notice that there is no difficulty in changing to a new h in mid-integration if the convergence either deteriorates or improves.

The Taylor series method, as described above, has advantages and disadvantages over other methods which will be discussed in this, and the next, chapter. In its favour, it is applicable to a very wide class of problems, including non-linear systems. Also, it requires no special starting procedure. On the other hand, the evaluation of the necessary derivatives may often be difficult. Furthermore, it is presently unreasonable to expect an automatic digital computer program to evaluate the derivatives of an arbitrary function, given only its form; that is, without formulae for these derivatives being worked out by hand. However, some progress has been made in this area. Also, for some special classes of f_i, recursion formulae may be available for the derivatives.

We notice also that the Taylor series method can be applied directly to second and higher order equations without first transforming them to sets of first order equations.

Section 3.3 Euler's Method for Inital Value Problems

Euler [7], devised a method for the approximate solution of systems like (3.2.7) and (3.2.6). Before describing it, we will change our notation to make it more compact.

We will write y_i of (3.2.6) and (3.2.7) as components of a vector

$$\underline{y}(x) = \begin{pmatrix} y_o(x) \\ y_1(x) \\ \vdots \\ y_n(x) \end{pmatrix} . \tag{3.3.1}$$

Hence, we can write

$$f_i(x,y_o,y_1\ldots,y_n) = f_i(x,\underline{y}). \tag{3.3.2}$$

Similarly, we write

$$\underline{f}(x,\underline{y}) = \begin{pmatrix} f_o(x,\underline{y}) \\ f_1(x,\underline{y}) \\ \vdots \\ f_n(x,\underline{y}) \end{pmatrix} , \tag{3.3.3}$$

and hence, we can now rewrite equations (3.2.7) as the

vector equation

$$\underline{y}'(x) = f(x,\underline{y}). \tag{3.3.4}$$

If we also write the initial conditions in vector form, as

$$\underline{\eta} = \begin{pmatrix} \eta_0 \\ \eta_1 \\ . \\ . \\ . \\ \eta_n \end{pmatrix} , \tag{3.3.5}$$

then (3.2.6) is compactly written

$$\underline{y}(x_0) = \underline{\eta} . \tag{3.3.6}$$

Operations on a vector, such as differentiation, differencing, integration, etc., are to be interpreted as acting element by element on the vector.

Euler's method can now be obtained by at least three methods:

(i) Consider the Taylor series method of Section 3.2. Using our new vector notation, equations (3.2.15) may be written

$$\underline{y}(x_0+h) = \underline{y}(x_0)+h\ \underline{y}'(x)+ \frac{h^2}{2!}\ \underline{y}''(x)+... \quad . \tag{3.3.7}$$

If we truncate this Taylor series after the first derivative term, and substitute (3.3.4) into it, we have

$$\underline{y}(x_o+h) \approx \underline{y}(x_o) + h\underline{f}(x_o,\underline{y}(x_o)). \qquad (3.3.8)$$

(ii) Alternatively, we may integrate (3.3.4) from x_o to x_o+h:

$$\int_{x_o}^{x_o+h} \underline{y}'(t)dt = \int_{x_o}^{x_o+h} \underline{f}(t,\underline{y}(t))dt. \qquad (3.3.9)$$

That is,

$$\underline{y}(x_o+h)-\underline{y}(x_o) = \int_{x_o}^{x_o+h} \underline{f}(t,\underline{y}(t))dt, \qquad (3.3.10)$$

and, approximating the integral on the right of (3.3.10) by the product of the interval of integration, and the value of the integrand at the left-hand end of the interval,

$$\underline{y}(x_o+h) - \underline{y}(x_o) \approx h\, \underline{f}(x_o,\underline{y}(x_o)), \qquad (3.3.11)$$

and (3.3.8) follows immediately.

(iii) Finally, we use the approximation (2.4.3),

$$h\, D \approx \Delta.$$

That is

$$D \approx \frac{\Delta}{h}, \qquad (3.3.12)$$

and thus,

$$\underline{y}'(x_o) = D\underline{y}(x_o) \approx \frac{1}{n}\Delta\underline{y}(x_o)$$

$$= \frac{\underline{y}(x_o+h)-\underline{y}(x_o)}{h} = \underline{f}(x_o,\underline{y}(x_o)). \qquad (3.3.13)$$

Equation (3.3.8) follows immediately from the last two members of (3.3.13).

The formula thus derived allows us to calculate an approximate value of $\underline{y}(x_o+h)$, given only $\underline{y}(x_o)$ and the formula for $\underline{f}$. In the same way as with the Taylor series method, we can advance the solution a step at a time until we reach the maximum desired value of x.

We exhibit this process as an algorithm:

→	1	$x \leftarrow x_o$	initialize x
	2	$\underline{y} \leftarrow \underline{\eta}$	initialize $\underline{y}$
	3	PRINT; x,$\underline{y}$	output solution point
	4	$\underline{y} \leftarrow \underline{y} + h\ \underline{f}(x,\underline{y})$	calculate next ordinate
	5	$x \leftarrow x+h$	calculate next abscissa
	6	$x : x_{max}, (\leq,>) \rightarrow (3,7)$	test for maximum abscissa
	7	STOP	

The importance of this algorithm is not so much practical, since, in practice, much more efficient algorithms are known, but is, rather, theoretical. It is the basis for an extremely important existence theorem, which we will discuss in the next section, and it will also serve to point out an ubiquitous difficulty in all initial value approximations - the problem of stability - to be taken up briefly in the section after next.

Section 3.4 Cauchy's Existence Theorem for Initial Value Problems

Subject to fairly mild restrictions on $\underline{f}(x,\underline{y})$, Cauchy [8] proved that a unique solution to (3.3.4) and (3.3.6) exists, and that, in the limit as $h \to 0$, the result of Euler's method converges to that solution. A detailed account of this proof is given by Henrici [9].

As a simple example, we consider the problem:

$$y'(x) = y(x) \qquad (3.4.1)$$

subject to the initial condition,

$$y(0) = 1. \qquad (3.4.2)$$

Let us suppose that we use Euler's method for n steps, each of length h, arriving finally at $x = nh$.

Then, according to (3.3.8),

$$y(x_o+h) = y(x_o) + hy(x_o) = (1+h)y(x_o), \tag{3.4.3}$$

and, in general,

$$\begin{aligned} y(0) &= 1, \\ y(h) &= (1+h), \\ y(2h) &= (1+h)^2, \\ &\vdots \\ y(nh) &= (1+h)^n \end{aligned} \tag{3.4.4}$$

Now, since $n = x/h$, the last of (3.4.4) gives

$$y(x) = (1+h)^{x/h} = \left[(1+h)^{1/h}\right]^x. \tag{3.4.5}$$

Now, as $h \to 0$, by the definition of e,

$$(1+h)^{1/h} \to e = 2.71828\ldots\ , \tag{3.4.6}$$

and so

$$y(x) \to e^x, \tag{3.4.7}$$

which is, indeed, the exact solution to the given problem. Thus, the convergence of Euler's approximation is made plausible at least.

Cauchy's theorem can be stated (and proved) in a strong form, and in a somewhat weaker form. The stronger

form, as proved in Henrici [9], follows.

Consider the system

$$\underline{y}'(x) = \underline{f}(x,\underline{y}), \qquad (3.3.4)$$

subject to

$$\underline{y}(x_o) = \underline{\eta}. \qquad (3.3.6)$$

Assume that $\underline{f}(x,\underline{y})$ satisfies the following two conditions:

(A) $\underline{f}(x,\underline{y})$ exists and is continuous for $x \in [a,b]$ and y_i, the components of $\underline{y}$ in $(-\infty,\infty)$, and, (B) there exists some constant L, called the Lipschitz constant, such that for $x \in [a,b]$ and for arbitarary vectors $\underline{y}$ and $\underline{y}^*$,

$$\| \underline{f}(x,\underline{y}) - \underline{f}(x,\underline{y}^*) \| \leq L \| \underline{y}-\underline{y}^* \| , \qquad (3.4.8)$$

where, for any vector, say $\underline{v}$, we define the "norm" of $\underline{v}$ as

$$\|\underline{v}\| = | v_1 | + | v_2 | + | v_3 | + \ldots \quad . \qquad (3.4.9)$$

Under assumptions (A) and (B), there exists a unique vector function $\underline{y}(x)$ which

(i) is continuous and everywhere differentiable at least once for $x \in [a,b]$,

(ii) which satisfies the differential system (3.3.4) everywhere for $x \in [a,b]$,

and (iii) which satisfies the initial conditions (3.3.6).

This theorem is stated and proved in a weaker form, namely for a neighbourhood only of $x = a$, in Tricomi [10].

Condition (B) is called a Lipschitz condition. It is a condition stronger than a requirement of continuity alone, but not as strong as requiring continuous differentiability.

Leaving the vector notation behind for the moment, and considering a single equation, we can readily see that (B) is satisfied by any continuously differentiable f, since, by the mean value theorem,

$$f(x,y) - f(x,y^*) = \frac{\partial f(x,\bar{y})}{\partial y} \cdot (y-y^*) \tag{3.4.10}$$

for some $\bar{y}$ in $[y,y^*]$.

If we consider, for example $f(x,y) = |y|$, a function which is not continuously differentiable, we find that it satisfies condition (B), since then

$$||f(x,y) - f(x,y^*)|| = |(|y| - |y^*|)| \leq |y-y^*|. \tag{3.4.11}$$

On the other hand, the function $f(x,y)=|y|^{1/2}$ is continuous, but does not satisfy a Lipschitz condition. It may be noted that for functions f not satisfying a Lipschitz condition, a solution may or may not exist or be unique. A theorem by Peano establishes the existence of at least one solution (i.e., not necessarily unique) if the Lipschitz condition is replaced by simple continuity [17].

Section 3.5 The Problem of Stability of Initial Value Problems

Cauchy's theorem of the previous section might lead us to believe that the solution of initial value problems is a fairly straightforward procedure. As we shall see in this section, such a belief is often quite unfounded.

Let us consider the second-order differential equation

$$y''(x) - y(x) = 0, \qquad (3.5.1)$$

subject to

$$y(0) = 1, \text{ and } y'(0) = -1. \qquad (3.5.2)$$

The exact solution of this system can readily be found to be

$$y(x) = e^{-x}. \qquad (3.5.3)$$

We will now develop a numerical method of approximating the solution to this system. The approach used here is rather atypical, but is used because it is somewhat revealing.

We first reduce (3.5.1) and (3.5.2) as described in Section 3.2 to

$$u'(x) = v(x)$$
$$v'(x) = u(x), \qquad (3.5.4)$$

subject to $u(0) = 1$ and $v(0) = -1$,
where $u(x) = y(x)$, and $v(x) = y'(x)$.

The Euler approximation gives us

$$u(x+h) = u(x) + hv(x) \qquad (3.5.5)$$
$$v(x+h) = v(x) + hu(x). \qquad (3.5.6)$$

Solving (3.5.5) for $v(x)$, we have

$$v(x) = \frac{u(x+h) - u(x)}{h}, \qquad (3.5.7)$$

and, substituting this into (3.5.6),

$$\frac{u(x+2h) - u(x+h)}{h} = \frac{u(x+h)-u(x)}{h} + hu(x). \qquad (3.5.8)$$

Hence, we obtain the second order difference equation

$$u(x+2h) - 2u(x+h) + (1-h^2)u(x) = 0, \tag{3.5.9}$$

as an approximation to (3.5.1).

Essentially, (3.5.9) allows us to calculate the next step in the solution, given the previous two. Now, the initial conditions (3.5.2) give us the value of $u(0)$. Let us suppose that, before we start, we use a Taylor series method, or some other method to obtain an accurate value of $u(h)$, or simply, suppose that $u(h)$ is given rather than $u'(0)$. In either case, let us use as the value of $u(h)$, e^{-h}, the exact value of the solution to the original problem.

The results of a numerical calculation using this method are given, for three values of h, in the third column of Table I, along with corresponding values of the exact solution, e^{-x}, in the second column.

The values obtained are drastically in error, expecially for $h = 0.5$. For $h = 0.1$, the approximation follows the correct answer fairly well for a while but eventually veers off and grows rapidly. We say that this approximation is "unstable."

TABLE I

SOLUTION OF

$$u(x+2h) = 2u(x+h) - (1-h^2)u(x)$$

x	e^{-x}	u(x) using $u(h) = e^{-h}$	u(x) using $u(h) = (1-h)$
		h = 0.5	
0.0	1.00000	1.00000	1.00000
0.5	.60653	.60653	.50000
1.0	.36788	.46306	.25000
1.5	.22313	.47123	.12500
2.0	.13534	.59516	.06250
2.5	.08209	.83689	.03125
3.0	.04979	1.22741	.01563
3.5	.03020	1.82716	.00781
4.0	.01832	2.73376	.00391
		h = 0.25	
0.0	1.00000	1.00000	1.00000
0.5	.60653	.62010	.56250
1.0	.36788	.43881	.31641
1.5	.22313	.38746	.17798
2.0	.13534	.43768	.10011
2.5	.08209	.58953	.05631
3.0	.04979	.86807	.03168
3.5	.03020	1.32651	.01782
4.0	.01832	2.05588	.01002
		h = 0.1	
0.0	1.00000	1.00000	1.00000
0.5	.60653	.61516	.59049
1.0	.36788	.40298	.34868
1.5	.22313	.30194	.20589
2.0	.13534	.28135	.12158
2.5	.08209	.33210	.07179
3.0	.04979	.46339	.04239
3.5	.03020	.70411	.02503
4.0	.01832	1.10906	.01478

To see why this is the case we will solve the difference equation (3.5.9) analytically. To do so, let us set

$$x = nh \text{ and } u(x) = u(nh) = u_n. \qquad (3.5.10)$$

Then we may write (3.5.9)

$$u_{n+2} - 2u_{n+1} + (1-h^2)u_n = 0. \qquad (3.5.11)$$

As a trial solution, let us use

$$u_n = C^n, \qquad (3.5.12)$$

for some constant, C.

Then

$$C^{n+2}-2C^{n+1}+(1-h^2)C^n=0, \qquad (3.5.13)$$

and so

$$C^2-2C+(1-h^2) = 0, \qquad (3.5.14)$$

of which the roots are

$$C = 1\pm h. \qquad (3.5.15)$$

Hence, we shall write

$$u_n = A(1+h)^n+B(1-h)^n \qquad (3.5.16)$$

To determine the coefficients A and B, we will

use the initial values $u(0) = 1$, $u(h) = e^{-h}$, to obtain; for $n = 0$,

$$1 = A+B \tag{3.5.17}$$

and for $n = 1$,

$$e^{-h} = A(1+h)+B(1-h). \tag{3.5.18}$$

From the first of these,

$$A = 1-B, \tag{3.5.19}$$

and therefore,

$$e^{-h} = (1-B)(1+h)+B(1-h). \tag{3.5.20}$$

Hence,

$$B = \frac{h+(1-e^{-h})}{2h}, \tag{3.5.21}$$

and therefore,

$$A = \frac{h-(1-e^{-h})}{2h} . \tag{3.5.22}$$

The source of the difficulty is now clear. The differential equation (3.5.1) has both an increasing and a decreasing solution, but the initial conditions specify a coefficient of zero for the increasing solution. The difference equation (3.5.12) also has increasing and decreasing solutions, $(1+h)^n$ and $(1-h)^n$, but the initial

conditions that we used allow a component of the increasing solution to remain. This component quite quickly dominates the decreasing solution, with the results which we have observed.

We can see that, in a more general situation, the difference equation might have several increasing, or several decreasing solutions, and that a component of an unwanted solution might creep in and swamp the desired solution.

It can be seen that if we had used $u(h) = 1-h$ instead of e^{-h}, that we would have obtained $A = 0$ and $B = 1$. In other words, the increasing solution would have been excluded. The fourth column of Table I lists solutions of (3.5.9) using $u(h) = 1-h$. We observe a very much improved behaviour, although the approximation is still far from excellent.

It is interesting to note that had we, in obtaining our value of $u(h)$ been less fastidious and used (3.5.5) instead of the true value $u(h) = e^{-h}$, we would have obtained $u(h) = 1-h$.

We have now illustrated the fact that the exact initial conditions of the differential equation may often (in fact, almost always) not be the most appropriate

conditions to use in a difference approximation to that differential equation.

Also worth noting is that, even using $u(h) = 1-h$, we expect round-off error in our arithmetic to introduce a small component of the increasing solution, and that, eventually, the decreasing solution would become dominated. However, this situation is far less serious than using $u(h) = e^{-h}$, as Table I shows.

The topic of stability is a large one, as we shall return to it briefly in later sections.

Section 3.6 Simple Approximations for Boundary Value Problems

In this section, we will consider linear differential equations subject to boundary conditions. Existence theorems are available for these systems [9],[10]. For instance [9], for the problem

$$y'' = f(x,y), \tag{3.6.1}$$

subject to

$$y(a) = \alpha, \; y(b) = \beta, \tag{3.6.2}$$

there exists a unique solution y, provided that $\frac{\partial f(x,y)}{\partial y}$ is continuous and non-negative for $x \in [a,b]$, and $y \in (-\infty,\infty)$.

Although most of the results of this section can readily be extended to higher-order equations, we will, for simplicity, consider only the following second-order equation:

$$y''(x) + f(x)y'(x) + g(x)y(x) = k(x). \tag{3.6.3}$$

We will first suppose that we have boundary conditions of the form (3.6.2). We will attempt to calculate approximate values y_i of $y(x)$ at the n+1 points

$$x = a,\ a+h,\ a+2h,\ldots,\ b-h,\ b; \tag{3.6.4}$$

we write

$$y_i \approx y(a+ih). \tag{3.6.5}$$

Let us begin by replacing the derivatives in (3.6.3) by the difference expansions (2.3.11) and (2.3.15), each truncated to a single term; thus

$$h^2D^2 \approx \delta^2, \tag{3.6.7}$$

and

$$hD \approx \mu\delta. \tag{3.6.8}$$

Then, we will have

$$h^2D^2y_i \approx \delta^2 y_i = y_{i+1}-2y_i+y_{i-1}, \qquad (3.6.10)$$

and

$$hDy_i \approx \mu\delta y_i = \frac{1}{2}(y_{i+1}-y_{i-1}), \qquad (3.6.11)$$

which, when substituted into (3.6.3), yield

$$(y_{i+1}-2y_i+y_{i-1}) + \frac{hf_i}{2}(y_{i+1}-y_{i-1}) + h^2g_iy_i = h^2k_i,$$

i.e.,

$$y_{i+1}(1+\frac{hf_i}{2}) + y_i(h^2g_i-2) + y_{i-1}(1-\frac{hf_i}{2}) = h^2k_i. \qquad (3.6.12)$$

Writing down (3.6.12) for successive values of i, and keeping the known boundary values in mind, we have:

$$\begin{aligned}
-(2-h^2g_1)y_1+(1+\tfrac{1}{2}hf_1)y_2 &= h^2k_1-\alpha(1-\tfrac{1}{2}hf_1)\\
(1-\tfrac{1}{2}hf_2)y_1-(2-h^2g_2)y_2+(1+\tfrac{1}{2}hf_2)y_3 &= h^2k_2\\
(1-\tfrac{1}{2}hf_3)y_2-(2-h^2g_3)y_3+(1+\tfrac{1}{2}hf_3)y_3 &= h^2k_3\\
\cdots\cdots\cdots

$$(1-\tfrac{1}{2}hf_{n-2})y_{n-3}-(2-h^2g_{n-2})y_{n-2}+(1+\tfrac{1}{2}hf_{n-2})y_{n-1} = h^2k_{n-2}$$

$$(1-\tfrac{1}{2}hf_{n-1})y_{n-2}-(2-h^2g_{n-1})y_{n-1} = h^2k_{n-1}-\beta(1+\tfrac{1}{2}hf_{n-1}),$$

a set of (n-1) linear equations in the (n-1) unknowns $y_1, y_2, \ldots, y_{n-1}$.

Let us write (3.6.13) in matrix-vector form:

$$\underline{A}\ \underline{y} = \underline{d}, \qquad (3.6.14)$$

where $\underline{A}$ is the (n-1) by (n-1) square matrix

$$\begin{pmatrix}
-(2-h^2g_1) & (1+\frac{hf_1}{2}) & 0 & \cdots \\
(1-\frac{hf_2}{2}) & -(2-h^2g_2) & (1+\frac{hf_2}{2}) & \cdots \\
0 & (1-\frac{hf_3}{2}) & -(2-h^2g_3) & (1+\frac{hf_3}{2}) \\
\cdots & \cdots & \cdots & \cdots \\
\cdots & (1-\frac{hf_{n-2}}{2}) & -(2-h^2g_{n-2}) & (1+\frac{hf_{n-2}}{2}) \\
\cdots & 0 & (1-\frac{hf_{n-1}}{2}) & -(2-h^2g_{n-1})
\end{pmatrix},$$

$\underline{y}$ is the column vector

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{pmatrix},$$

and $\underline{d}$ is the column vector

$$\begin{pmatrix} h^2k_1 - \alpha\left(1 - \frac{hf_1}{2}\right) \\ h^2k_2 \\ h^2k_3 \\ \vdots \\ h^2k_{n-2} \\ h^2k_{n-1} - \beta\left(1 + \frac{hf_{n-1}}{2}\right) \end{pmatrix}.$$

Since the matrix $\underline{A}$ is triple-diagonal, (3.6.14) may most readily be solved by the LU factorization method [11]. When this method is programmed for a digital computer, we note that storage requirements go up only linearly (rather than quadratically) and the computing time also goes up linearly (rather than cubically) with the number of equations to be solved.

Suppose that, instead of boundary conditions of the form (3.6.2), we are given

$$y'(a) + \gamma y(a) = \alpha \qquad (3.6.15)$$

in place of $y(a) = \alpha$.

Then, using (3.6.12) with $i = 0$, we have

$$(1- \frac{hf_o}{2})y_{-1}-(2-h^2g_o)y_o+(1+ \frac{hf_o}{2})y_1 = h^2k_o \qquad (3.6.16)$$

and the finite difference replacement for (3.6.15) is

$$y_1-y_{-1}+2h\gamma y_o = 2h\alpha. \qquad (3.6.17)$$

Eliminating y_{-1} between (3.6.16) and (3.6.17), we obtain

$$[(1- \frac{hf_o}{2})2h\gamma-(2-h^2g_o)]y_o+2y_1 = h^2k_o+2h\alpha(1- \frac{hf_o}{2}). \qquad (3.6.18)$$

This becomes an additional first equation of the set (3.6.13), and the second equation (replacing the old first equation) is clearly

$$(1- \frac{hf_1}{2})y_o-(2-h^2g_1)y_1+(1+ \frac{hf_1}{2})y_2 = h^2k_1. \qquad (3.6.19)$$

The remaining members of (3.6.13) are unaltered unless the boundary condition at x = b also contains a derivative, in which case (3.6.13) acquires an additional last equation, and the old last equation is modified in the obvious way.

Finally, we will consider briefly a class of eigenvalu differential equations. If we are given

$$y''(x)+f(x)y'(x)+g(x)y(x) = \lambda y(x), \quad (3.6.20)$$

subject to

$$y(a) = y(b) = 0, \quad (3.6.21)$$

then we can perform the analysis of the beginning of this section. When we do so, we obtain a system of equations like (3.6.13) or (3.6.14) except for the right hand side $\underline{d}$. Clearly, $\underline{d}$ is replaced by $h^2\lambda\underline{y}$.

Thus, we have to solve the problem

$$\underline{A}\ \underline{y} = h^2\lambda\underline{y}\ , \quad (3.6.22)$$

or, setting

$$\mu = h^2\lambda \quad (3.6.23)$$

we have to solve

$$\underline{A}\ \underline{y} = \mu\underline{y} \quad (3.6.24)$$

Equation (3.6.24) is the standard algebraic eigenproblem, and may be solved by standard techniques which need not concern us here.

Section 3.7 Classification of Second-Order Partial Differential Equations

We will consider the general, second-order quasi-linear partial differential equation:

$$a \frac{\partial^2 u}{\partial x^2} + b \frac{\partial^2 u}{\partial x \partial y} + c \frac{\partial^2 u}{\partial y^2} = e, \qquad (3.7.1)$$

where a,b,c, and e are real functions of $u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, x$, and y, but not of the second derivatives (hence the classification "quasi-linear").

The following discussion could equally well be carried out with respect to a pair of first-order equations equivalent to (3.7.1), instead of the single second-order equation.

A standard notation to compact (3.7.1) is

$$p = \frac{\partial u}{\partial x}, \quad q = \frac{\partial u}{\partial y}, \quad r = \frac{\partial^2 u}{\partial x^2}, \quad s = \frac{\partial^2 u}{\partial x \partial y}, \quad t = \frac{\partial^2 u}{\partial y^2}, \qquad (3.7.2)$$

whence (3.7.1) becomes

$$ar + bs + ct = e. \qquad (3.7.3)$$

The classification of (3.7.3) is motivated by a simple attempt to solve it. Let us suppose that we are given starting data in the form of values of u,p, and q along some curve, say L, in the (x,y) plane described by a parameter ℓ. That is to say, we are given u, p, and q as

$$u = u(\ell),\ p = p(\ell),\ \text{and}\ q = q(\ell), \tag{3.7.4}$$

on the curve L defined by

$$x = x(\ell)\ \text{and}\ y = y(\ell). \tag{3.7.5}$$

Our attempted solution is by means of a generalization of the Taylor series method of Section 3.2. Our starting conditions give us the value of u and its two first derivatives at any point on L. Using this information, and the differential equation (3.7.3), let us attempt to find values for the three second derivatives, four third derivatives and so on, of u.

By simple calculus, we can write

$$dp = \frac{\partial p}{\partial x}dx + \frac{\partial p}{\partial y}\,dy, \tag{3.7.6}$$

or

$$dp = rdx + sdy. \tag{3.7.7}$$

If we differentiate along L, i.e., with respect to ℓ, then

$$\frac{dp}{d\ell} = r \frac{dx}{d\ell} + s \frac{dy}{d\ell} . \qquad (3.7.8)$$

Similarly, by differentiating q along L,

$$\frac{dq}{d\ell} = s \frac{dx}{d\ell} + t \frac{dy}{d\ell} \qquad (3.7.9)$$

Thus, (3.7.3), (3.7.8), and (3.7.9) are three equations in the three unknowns r, s, and t which we seek. We write these three equations in matrix form as

$$\begin{pmatrix} a & b & c \\ \frac{dx}{d\ell} & \frac{dy}{d\ell} & 0 \\ 0 & \frac{dx}{d\ell} & \frac{dy}{d\ell} \end{pmatrix} \begin{pmatrix} r \\ s \\ t \end{pmatrix} = \begin{pmatrix} e \\ \frac{dp}{d\ell} \\ \frac{dq}{d\ell} \end{pmatrix} , \qquad (3.7.10)$$

or

$$\underline{A}\ \underline{f} = \underline{g} . \qquad (3.7.11)$$

If we suppose that we have solved (3.7.11) for r, s, and t, then we can differentiate (3.7.11) with respect, say, to x, to obtain

$$\underline{A} \frac{\partial \underline{f}}{\partial x} + \frac{\partial \underline{A}}{\partial x} \underline{f} = \frac{\partial \underline{g}}{\partial x} \qquad (3.7.12)$$

i.e.,

$$\underline{A}\,\frac{\partial \underline{f}}{\partial x} = \frac{\partial \underline{g}}{\partial x} - \frac{\partial \underline{A}}{\partial x}\,\underline{f}\ , \qquad (3.7.13)$$

and we then solve (3.8.12) for $\frac{\partial \underline{f}}{\partial x}$.

If we have already obtained $\underline{A}^{-1}$ in solving (3.7.11), then we may use it again, and write

$$\frac{\partial \underline{f}}{\partial x} = \underline{A}^{-1}\left[\frac{\partial \underline{g}}{\partial x} - \frac{\partial \underline{A}}{\partial x}\,\underline{f}\right] . \qquad (3.7.14)$$

The elements of $\frac{\partial \underline{f}}{\partial x}$ are, obviously,

$$\frac{\partial r}{\partial x} = \frac{\partial^3 u}{\partial x^3};\ \frac{\partial s}{\partial x} = \frac{\partial^3 u}{\partial x^2 \partial y}\ ;\ \frac{\partial t}{\partial x} = \frac{\partial^3 u}{\partial x \partial y^2}\ , \qquad (3.7.15)$$

three of the required four third derivatives of u.

The fourth can be found by differentiating (3.7.11) with respect to y, and proceeding as above.

In this fashion we can, at least in principle, go on to obtain derivatives of all orders, and carry out our Taylor series scheme.

Unfortunately, there is an implicit assumption involved here, namely that the matrix $\underline{A}$ is non-singular.

Since the manifold L has so far been arbitrary, and since a, b, and c are arbitrary, there is no reason why $\underline{A}$ should not be singular.

If, indeed, $\underline{A}$ is singular, then

$$|\underline{A}| = a\left(\frac{dy}{d\ell}\right)^2 - b\frac{dx}{d\ell}\frac{dy}{d\ell} + c\left(\frac{dx}{d\ell}\right)^2 = 0, \qquad (3.7.16)$$

which we rewrite as

$$a\left(\frac{dy}{dx}\right)^2 - b\frac{dy}{dx} + c = 0, \qquad (3.7.17)$$

or, finally, as

$$\frac{dy}{dx} = \frac{b \pm \sqrt{b^2-4ac}}{2a} . \qquad (3.7.18)$$

Clearly, if $\Delta = b^2-4ac$ is positive, then (3.7.18) defines, at any given point on L, two directions along which the Taylor series scheme is certain to fail. If $\Delta = 0$, there is only one such direction, and if $\Delta<0$, there are no such directions in the real domain.

According as $\Delta>0$, $\Delta=0$, or $\Delta<0$, we term the differential equation (3.7.3) "hyperbolic," "parabolic," or "elliptic," respectively [12].

It is interesting to notice that, since, in general, a, b, and c may depend upon u, the character

(i.e., hyperbolic, etc.) of a given equation may change from region to region.

Either of (3.7.17) or (3.7.18) is called the "characteristic equation" of (3.7.3) and the solutions are called "characteristics" or "jump lines." The reason for the latter term is clear, since the jump lines are lines along which the Taylor series for u may fail to exist, and hence, along which u may be discontinuous. This property is elucidated in some detail in [13], page 416.

It can be shown that the type of a partial differential equation determines to a very great extent the type of boundary conditions which are required to make the equation solvable in a realistic manner. A discussion of this topic may be found in [12], or, at greater length, in [13].

We will discuss typical methods for solving systems of the three types, in the succeeding three sections.

Section 3.8 The Method of Characteristics for Hyperbolic Systems

The characteristic curves described in the previous section prevented our easy exploitation of the

Taylor series approach to a solution. In this section, we make use of these same characteristics to obtain an approximate solution to a hyperbolic equation of the form (3.7.3). The notation and exposition of this section follow that of [11].

Referring to Figure I, let us suppose that we are given the values of u, p, and q on a segment, AB, of some non-characteristic curve. Considering any two points P and Q on AB (which are fairly close together), then two characteristics pass through each of P and Q. These curves satisfy

$$\frac{dy}{dx} = \frac{b \pm \sqrt{b^2-4ac}}{2a} \quad , \qquad (3.7.18)$$

and we shall distinguish the two cases by writing them as

$$\frac{dy}{dx} = f, \text{ and } \frac{dy}{dx} = g, \qquad (3.8.1)$$

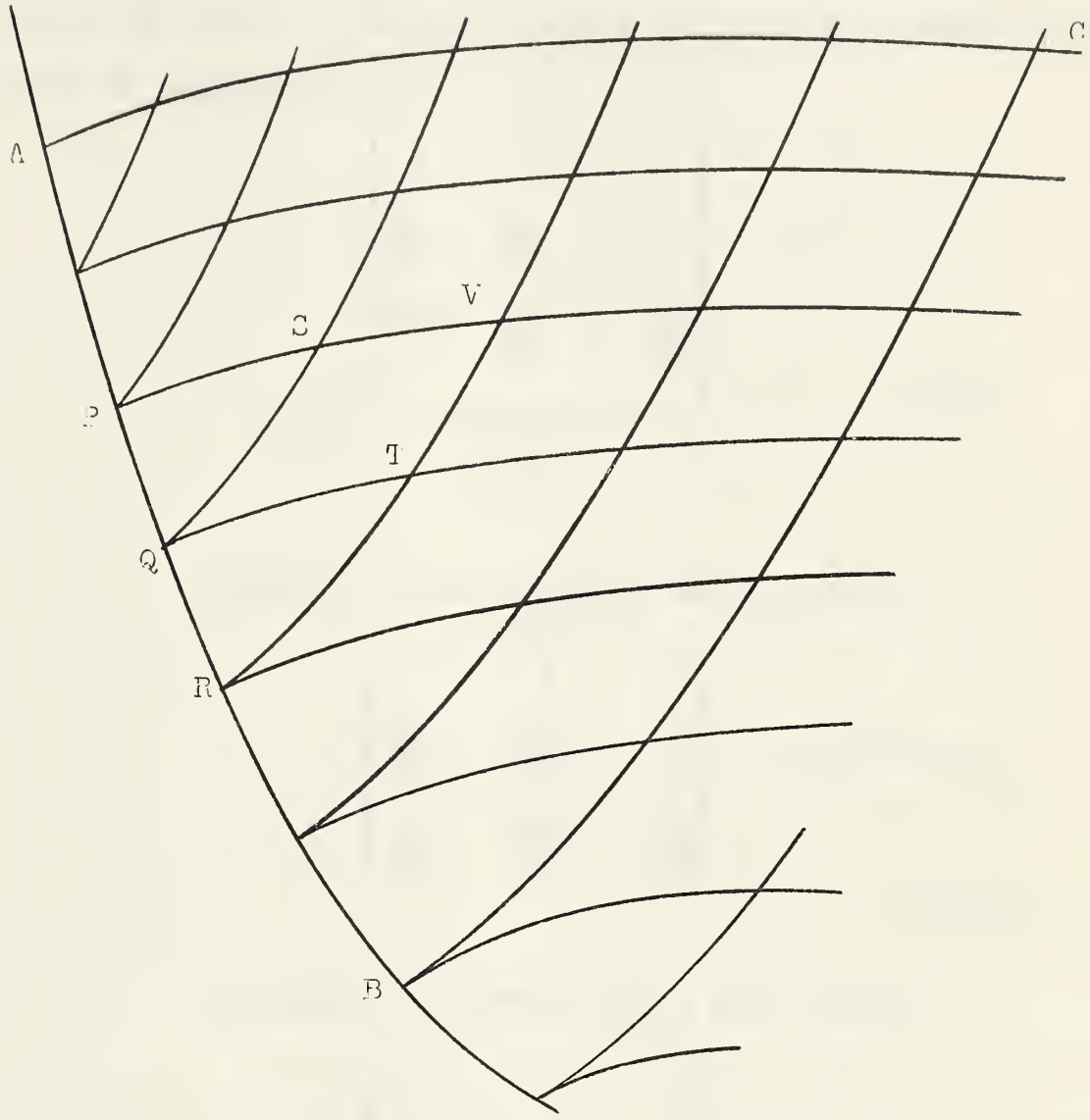
where

$$f = \frac{b+\sqrt{b^2-4ac}}{2a} \quad , \text{ and } g = \frac{b-\sqrt{b^2-4ac}}{2a} \quad . \qquad (3.8.2)$$

Returning now to (3.7.11), since we are on a characteristic, $|\underline{A}| = 0$. Since we are seeking an analytic

FIGURE I

THE CHARACTERISTIC CURVES OF A HYPERBOLIC PARTIAL DIFFERENTIAL EQUATION


A
C
V
T
Q
R
B

solution, it is clear that for such a solution to be possible, the quantities r, s, and t must none-the-less exist. That is if we were to attempt to solve (3.7.11) by Cramer's rule, then we would have for r,

$$r = \frac{\begin{vmatrix} e & b & c \\ \frac{dp}{d\ell} & \frac{dy}{d\ell} & 0 \\ \frac{dq}{d\ell} & \frac{dx}{d\ell} & \frac{dy}{d\ell} \end{vmatrix}}{|\underline{A}|} . \qquad (3.8.3$$

Hence, if r is to exist, we must have

$$\begin{vmatrix} e & b & c \\ \frac{dp}{d\ell} & \frac{dy}{d\ell} & 0 \\ \frac{dq}{d\ell} & \frac{dx}{d\ell} & \frac{dy}{d\ell} \end{vmatrix} = 0, \text{ also.}$$

(3.8.4)

Similarly, in order that s and t exist,

$$\begin{vmatrix} a & e & c \\ \frac{dx}{d\ell} & \frac{dp}{d\ell} & 0 \\ 0 & \frac{dq}{d\ell} & \frac{dy}{d\ell} \end{vmatrix} = 0,$$

(3.8.5)

and

$$\begin{vmatrix} a & b & e \\ \frac{dx}{d\ell} & \frac{dy}{d\ell} & \frac{dp}{d\ell} \\ 0 & \frac{dx}{d\ell} & \frac{dq}{d\ell} \end{vmatrix} = 0. \qquad (3.8.6)$$

The previous three equations are equivalent on a characteristic and we arbitrarily choose (3.8.5), to give

$$e \frac{dx}{d\ell} \frac{dy}{d\ell} - a \frac{dp}{d\ell} \frac{dy}{d\ell} - c \frac{dx}{d\ell} \frac{dq}{d\ell} = 0, \qquad (3.8.7)$$

which we rewrite as

$$e\, dy - a\, dp \frac{dy}{dx} - c\, d\, q = 0. \qquad (3.8.8)$$

Let the characteristic of the f type through P meet the characteristic of the g type through Q at some point S. Then, to a first approximation, we regard PS as a straight line with slope f_P. Hence, approximately,

$$y_S - y_P = f_P(x_S - x_P), \qquad (3.8.9)$$

and, by similarly regarding QS as straight, with slope g_Q,

$$y_S - y_Q = g_Q(x_S - s_Q). \qquad (3.8.10)$$

Equations (3.8.9) and (3.8.10) are two linear equations which we may readily solve for x_S and y_S, since all the other quantities are known.

Hence, using (3.8.8) along the lines PS and QS, and using our initial estimates of x_S and y_S, we can write

$$e_P(y_S-y_P)-a_P(p_S-p_P)f_P-c_P(q_S-q_P) = 0, \tag{3.8.11}$$

and

$$e_Q(y_S-y_Q)-a_Q(p_S-p_Q)g_Q-c_Q(q_S-q_Q)= 0. \tag{3.8.12}$$

Again, (3.8.11) and (3.8.12) are two linear equations which we can solve for p_S and q_S. Finally, we may obtain u_S approximately from

$$u_S = u_P+\left[\frac{\partial u}{\partial x}\right] + \left[\frac{\partial u}{\partial y}\right]dy, \tag{3.8.13}$$

where by $\left[\frac{\partial u}{\partial x}\right]$ and $\left[\frac{\partial u}{\partial y}\right]$ we intend mean values along PS. Hence,

$$u_S = u_P+\left[\frac{p_S+p_P}{2}\right](x_S-x_P)+\left[\frac{q_S+q_P}{2}\right](y_S-y_P) \tag{3.8.14}$$

Having now obtained an approximate value of u at S, we can improve it by the following process:

(i) Substitute the approximate values of u_S, p_S, and q_S into (3.8.2) to obtain approximate values of f_S and g_S.

(ii) In place of (3.8.9) and (3.8.10), use

$$y_S - y_P = \left(\frac{f_S + f_P}{2}\right)(x_s - x_P), \tag{3.8.16}$$

and

$$y_S - y_Q = \left(\frac{g_S + g_Q}{2}\right)(x_S - x_Q), \tag{3.8.17}$$

to obtain improved values of x_S and y_S.

(iii) In place of (3.8.11) and (3.8.12), use

$$\left[\frac{e_P + e_S}{2}\right](y_S - y_P) - \left[\frac{a_P + a_S}{2}\right](p_S - p_P)\left[\frac{f_P + f_S}{2}\right] - \left[\frac{C_P + C_S}{2}\right](q_S - q_P) = 0, \tag{3.8.18}$$

and

$$\left[\frac{e_Q + e_S}{2}\right](y_S - y_Q) - \left[\frac{a_Q + a_S}{2}\right](p_S - p_Q)\left[\frac{f_Q + f_S}{2}\right] - \left[\frac{C_Q + C_S}{2}\right](q_S - q_Q) = 0, \tag{3.8.19}$$

to obtain improved values of p_S and q_S.

(iv) Use (3.8.14) to obtain an improved u_S.

Steps (i) to (iv) are to be repeated until two successive values of u_s agree to the desired accuracy.

In a similar fashion, using the points Q and R in Figure I, we can obtain u_T and its derivatives. Then building out from S and T, we obtain the solution at V.

Finally, we can build up a complete solution in the region ABC. It is clear from our method of solution that we can go no further, because points outside of ABC depend on values of u, p, and q on the continuation of AB. Furthermore, it is clear that the solution inside of ABC is entirely independent of such values.

Section 3.9 Solution of Parabolic Systems

In this section, we will only illustrate the solution of parabolic equations by consideration of a specific parabolic equation, the dimensionless form of the heat equation:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial f}{\partial t} . \qquad (3.9.1)$$

As subsidiary conditions, f is specified on three sides of a rectangle, say $t = 0$, and $x = \pm 1$.

Probably the simplest procedure for parabolic equations is to construct a finite difference approximation to the second derivative term, thereby generating a set of initial value equations.

Dividing the range [-1,1] into equal intervals of width h, with tabular points $x_o, x_1 \ldots, x_n$, and denoting f evaluated as a function of t for $x = x_i$ by $f_i(t)$, we can replace (3.9.1) by the equations

$$h^2 \frac{df_i(t)}{dt} = f_{i-1} - 2f_i + f_{i+1}; \quad i=1,2,\ldots,n-1 \tag{3.9.2}$$

where the right hand side is the simple δ^2 approximation to the second derivative.

Writing down (3.9.2) for all values of i, and including the boundary conditions, we have

$$\begin{aligned}
h^2 \frac{df_1}{dt} + 2f_1 - f_2 &= f_o \\
h^2 \frac{df_2}{dt} - f_1 + 2f_2 - f_3 &= 0 \\
h^2 \frac{df_3}{dt} - f_2 + 2f_3 - f_4 &= 0 \qquad (3.9.3) \\
\cdots\cdots\cdots

The subsidiary conditions provide all the f_r at t=0, and f_o and f_n for all t. Hence (3.9.3) is a system of the type of (3.2.7) and (3.2.6) and is solvable by the techniques of Sections 3.2 and 3.3, or by the more sophistocated techniques described in the next chapter. It is also subject, of course, to the same stability problems that we saw in Section 3.5.

Section 3.10 Solution of Elliptic Systems

Boundary conditions for elliptic equations must be given on a closed contour. The usual method of solution is to divide the region inside the contour into a grid of square or rectangular cells, replace the differential equation by finite difference equations, and solve the resulting algebraic equations.

By way of illustration, we will consider the Laplace equation,

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0, \qquad (3.10.1)$$

with values of f given on a rectangular boundary.

Referring to a particular point (x_o, y_o) of the grid, we may write

$$(\delta x)^2 \frac{\partial^2 f_o}{\partial x^2} \approx \delta_x^2 f_o, \qquad (3.10.2)$$

where the subscript x on δ_x implies differencing in the x-direction. A similar expression can be written for the derivative with respect to y, whence we will write

$$\delta_x^2 f_o = (f_1 - 2f_o + f_{-1})_x \; ; \; \delta_y^2 f_o = (f_1 - 2f_o + f_{-1})_y . \tag{3.10.3}$$

We thus replace (3.10.1) by

$$(f_1 - 2f_o + f_{-1})_x + \left(\frac{\delta x}{\delta y}\right)^2 (f_1 - 2f_o + f_{-1})_y = 0, \tag{3.10.4}$$

which, in the case $\delta x = \delta y$, becomes

$$(f_1 + f_{-1})_x + (f_1 + f_{-1})_y - 4f_o = 0 \tag{3.10.5}$$

Thus, we can set up an equation like (3.10.5) at every point of the grid. If we have a hundred subdivisions in the x and the y direction, then we will have ten thousand simultaneous linear equations, in as many unknowns, to solve. Practical techniques for solving such huge sets of equations on a digital computer are described by Varga [14].

A technique for handling non-rectangular

boundary contours is described by Murray [15]. Detailed discussions of these and many other problems associated with partial differential equations may be found in [12], [13], and [19].

CHAPTER IV

ADVANCED NUMERICAL METHODS FOR ORDINARY DIFFERENTIAL EQUATIONS

In this chapter we develop and discuss practical methods for the approximate solutions of ordinary differential equations.

General multi-step methods are discussed, including forward integration and predictor-corrector formulas. The method of Milne is developed in detail, and an algorithm for it is presented.

The stability of multi-step methods is analyzed in some detail. Milne's method is shown to be unstable, but Hamming's modification to Milne's method is presented and shown to be conditionally stable.

Runge-Kutta methods are introduced, and a second order formula is developed. An algorithm for the Runge-Kutta-Gill fourth order process is given.

Difference correction methods and a method of linear combinations of initial value solutions for boundary value problems are outlined.

Section 4.1 Introduction to Multi-step Methods

We consider in this section a large and very general class of numerical formulas for initial value problems - the multi-step methods. The problem to be solved is again

$$\underline{y}'(x) = \underline{f}(x,\underline{y}), \qquad (3.3.4)$$

subject to

$$\underline{y}(x_o) = \underline{\eta}. \qquad (3.3.6)$$

For simplicity of presentation, we will consider only a single first order equation, instead of a set of such equations. All the results of this section can be extended in a trivial fashion to sets, however. Thus, we consider

$$y'(x) = f(x,y), \qquad (4.1.1)$$

subject to

$$y(x_o) = y_o. \qquad (4.1.2)$$

The name multi-step arises from the form of the formulas - the approximate value y_n of $y(x_n)=y(x_o+nh)$ is expressed as a linear combination of y_i and y_i' for several values of $i \leq n$. More specifically, we can consider two types of multi-step formula, the "forward integration" and the "iterative" types:

(i) In forward integration formulas, y_{n+1} is expressed as a linear combination of y_i and y_i' for several values of $i \leq n$;

(ii) In iterative formulas, y_{n+1}' is included in the linear combination.

In general then, we will write

$$y_{n+1}=a_n y_n+a_{n-1}y_{n-1}+\ldots+a_{n-p}y_{n-p}$$
$$+h(b_{n+1}y'_{n+1}+b_n y'_n+\ldots+b_{n-p}y'_{n-p})$$
$$+E_n, \qquad (4.1.3)$$

where $b_{n+1}=0$ for a forward integration formula. E_n represents the error at the (n+1)th step, supposing that all the y_i and y'_i on the right in (4.1.3) are known exactly; (4.1.3) is to be regarded as an equation for determining y_{n+1} explicitly or implicitly.

The error, E_n, arises from two sources:

(i) Sincethe right-hand side of (4.1.3) is not an infinite sum of terms, it cannot in general represent y_{n+1} exactly. That is, (4.1.3) may be thought of as being obtained by truncating some infinite series. This type of error is therefore called "truncation error".

(ii) In any practical calculating device, the results of arithmetic operations must be rounded to a finite number of decimal or binary places. Thus, we are concerned also with "round-off error."

Obviously, the y_i and y'_i on the right of (4.1.3) will <u>not</u> be known exactly, since they themselves will in general have been calculated by (4.1.3) at an earlier stage, and will suffer from truncation and round-off errors. The study of the accumulation of errors assumes

crucial importance as, we shall see later. We have already had a taste of the problem in Section 3.5.

We begin the derivation of formulas by obtaining two forms of Newton's interpolation formula in terms of backward differences. Regarding p as continuously variable, we write

$$f_{n+p} = E^p f_n . \tag{4.1.4}$$

Now, by definition,

$$\nabla = 1-E^{-1}$$

and therefore,

$$E = (1-\nabla)^{-1}, \tag{4.1.5}$$

whence

$$E^p = (1-\nabla)^{-p} \tag{4.1.6}$$

$$= 1+p\nabla+\frac{p(p+1)}{2!}\nabla^2 + \frac{p(p+1)(p+2)}{3!}\nabla^3+\ldots . \tag{4.1.7}$$

Also,

$$E^p = E(E^{p-1}), \tag{4.1.8}$$

and thus

$$E^p=E\left[1+(p-1)\nabla+\frac{(p-1)(p)}{2!}\nabla^2+\ldots\right]. \tag{4.1.9}$$

Thus, applying (4.1.7) and (4.1.9) to y'_{n+p},

we obtain

$$y'_{n+p} = y'_n + p\nabla y'_n + \frac{p(p+1)}{2!}\nabla^2 y'_n + \ldots$$

$$+ \frac{p(p+1)\ldots(p+k-1)}{k!}\nabla^k y'_n + E_p, \qquad (4.1.10)$$

where

$$E_p = h^{k+1}\frac{p(p+1)\ldots(p+k)}{(k+1)!}D^{k+2}y(\eta); \qquad (4.1.11)$$

and also,

$$y'_{n+p} = y'_{n+1} + (p-1)\nabla y'_{n+1} + \frac{(p-1)p}{2!}\nabla^2 y'_{n+1} + \ldots$$

$$+ \frac{(p-1)(p)(p+1)\ldots(p+k-2)}{k!}\nabla^k y'_{n+1} + E^*_p, \qquad (4.1.12)$$

where

$$E^*_p = h^{k+1}\frac{(p-1)p(p+1)\ldots(p+k-1)}{(k+1)!}D^{k+2}y(\eta^*), \qquad (4.1.13)$$

and where both η and η^* are in $[x_o, x_n+ph]$.

Numerical integration formulae are obtained by substituting (4.1.10) or (4.1.12) into the relation

$$y_{n+1} = y_{n-r} + h\int_{-r}^{1} y'_{n+p}\,dp, \qquad (4.1.14)$$

to obtain, respectively, forward integration or iterative formulae.

For example, putting (4.1.10) into (4.1.14) with r = 0, we have

$$y_{n+1}=y_n+h(1+\frac{1}{2}\nabla+\frac{5}{12}\nabla^2+\ldots)y'_n+En, \qquad (4.1.15)$$

which is the Adams' method. Taking the special case where all the differences are discarded, we have

$$y_{n+1}=y_n+hy'_n+h^2y''(\eta), \qquad (4.1.16)$$

where $\eta \in (x_n, x_{n+1})$. This is yet another derivation of Euler's method, this time including an error term.

Rather remarkably, if we substitute (4.1.10) into (4.1.14) and keep r'th differences for r odd, the coefficient of the r'th difference is zero [16], and so r-1 or r^{th} differences provide the same accuracy. For r=1,3,5, we have, with all difference terms expanded,

$$y_{n+1}=y_{n-1}+2hy'_n+\frac{h^3}{3}y'''(\eta), \qquad (4.1.17)$$

$$y_{n+1}=y_{n-3}+\frac{4}{3}h(2y'_n-y'_{n-1}+2y'_{n-2})+\frac{14}{45}h^5y^{v}(\eta), \qquad (4.1.18)$$

and

$$y_{n+1}=y_{n-5}+\frac{3h}{10}(11y'_n-14y'_{n-1}+26y'_{n-2}-14y'_{n-3}$$

$$+11y'_{n-4})+\frac{41}{140}h^7y^{vii}(\eta), \qquad (4.1.19)$$

where $\eta \in (x_{n-r}, x_{n+1})$.

Note that none of these is self-starting, since they all depend upon data at least two steps back.

Now, putting (4.1.12) into (4.1.14), with r = 0, and keeping only one difference, we have

$$y_{n+1} = y_n + \frac{h}{2}(y'_{n+1} + y'_n) - \frac{h^3}{12} y'''(\eta), \qquad (4.1.20)$$

with $\eta \in (x_n, x_{n+1})$.

Obviously, we cannot employ (4.1.20) directly, since we have no value for y'_{n+1} available. We may, however proceed in the following way:

(i) Obtain some rough estimate $\bar{y}_{n+1}$ of the value of y_{n+1}. For instance, (4.1.16) or (4.1.17) might be used to obtain an estimate.

(ii) Using $\bar{y}_{n+1}$ directly in the differential equation, (4.1.1), obtain an approximate value $\bar{y}'_{n+1}$ of y'_{n+1}.

(iii) Using (4.1.20), with the estimate $\bar{y}'_{n+1}$, obtain a more accurate value of $\bar{y}_{n+1}$.

(iv) Repeat steps (ii) and (iii) until two successive values of $\bar{y}_{n+1}$ agree to the desired accuracy.

Usually, one would choose h small enough to provide convergence after very few iterations.

The use of one approximation (in this case (4.1.17), called a "predictor" followed by iterated use of another (in this case (4.1.20)), called a "corrector" is generally called a "predictor-corrector" technique.

The formulas (4.1.17) to (4.1.18) are predictors, of the second, fourth, and sixth "order" respectively. The "order" of any of these formulae is defined to be one less than the power of h in the error term.

Corresponding to (4.1.18) and (4.1.19) we may obtain the corresponding fourth and sixth order correctors. Again substituting (4.1.12) into (4.1.14), it can readily be shown that, for odd r, the coefficient of the (r+2)th difference is zero, and that we need therefore consider only (r+1) differences. For r=1 and 3, respectively, we have

$$y_{n+1}=y_{n-1}+\frac{h}{3}(y'_{n+1}+4y'_n+y'_{n-1})-\frac{h^5}{90}\,y^{v}(\eta), \qquad (4.1.21)$$

and

$$y_{n+1}=y_{n-3}+\frac{2h}{45}(7y'_{n+1}+32y'_n+12y'_{n-1}$$
$$+32y'_{n-2}+7y'_{n-3})-\frac{8}{945}\,h^7y^{vii}(\eta), \qquad (4.1.22)$$

where η is in (x_{n-1},x_{n+1}) and (x_{n-3},x_{n+1}), respectively.

Note that the numerical coefficients of the error terms of (4.1.21) and (4.1.22) are considerably smaller than those of (4.1.18) and (4.1.19), respectively. This is the basic motivation for using a predictor-corrector technique rather than a predictor alone.

An initial, obvious, disadvantage of the above forward integration, and predictor-corrector techniques, is that they are not self-starting. That is, at the beginning, to obtain y_1, we need not only y_o and y_o' but also $y_{-1}, y_{-2} \ldots$, and $y_{-1}', y_{-2}', \ldots$. In practice, we could use a Taylor series, Euler's method with very small h, or a Runge-Kutta method (to be described in Section 4.4) to get the solution going.

We discuss the practical considerations of a particular predictor-corrector procedure, Milne's Method, in the next section.

Section 4.2 Milne's Method

Equations (4.1.18) and (4.1.21) form the basis of Milne's method [18]. For completeness, we write

Predictor:

$$\bar{y}_{n+1} = y_{n-3} + \frac{4h}{3}(2y_n' - y_{n-1}' + 2y_{n-2}'), \qquad (4.2.1)$$

and

$$\bar{y}'_{n+1}=f(x_{n+1},\bar{y}_{n+1}); \qquad (4.2.2)$$

Corrector:

$$y_{n+1}=y_{n-1}+\frac{h}{3}(y'_{n+1}+4y'_n+y'_{n-1}). \qquad (4.2.3)$$

The predictor and corrector are to be used as described in the previous section; that is, approximate values $\bar{y}_{n+1}$ and $\bar{y}'_{n+1}$ are obtained from (4.2.1) and (4.2.2) and then successively improved by repeated use of (4.2.3) and (4.2.2).

The truncation error at each step can quite easily be estimated. According to (4.1.18), the truncation error of (4.2.1) is

$$T_1=\frac{14}{45}h^5y^{v}(\eta_1), \qquad (4.2.4)$$

for $\eta_1 \in(x_{n-3},x_{n+1})$. Similarly, the truncation error of (4.2.3) is

$$T_2= -\frac{1}{90}h^5y^{v}(\eta_2), \qquad (4.2.5)$$

for $\eta_2 \in(x_{n-1},x_{n+1})$.

Assuming the continuity of $y^{v}(x)$, we may thus

write

$$y_{n+1}-\bar{y}_{n+1}=\frac{1}{90}h^5y^v(\eta_2)+\frac{14}{45}h^5y^v(\eta_1)$$

$$=\frac{29}{90}h^5y^v(\eta), \qquad (4.2.6)$$

for some $\eta \in (x_{n-3},x_{n+1})$. If we further assume that $y^v(x)$ is relatively constant in (x_{n-3},x_{n+1}), then approximately,

$$\left.\begin{aligned} y_{n+1}-\bar{y}_{n+1}&=\frac{29}{90}h^5y^v(\eta)\approx\frac{29}{28}T_1 \\ &\approx -29T_2. \end{aligned}\right\} \qquad (4.2.7)$$

Thus, by calculating $y_{n+1}-\bar{y}_{n+1}$ at each step we obtain a good indication of the local truncation error. If this quantity grows smaller over several successive steps, then we may increase h, or, if it grows larger, we must decrease h.

One major difficulty with the method described above is that, in the process of iterating, we must evaluate the function f each time around. Thinking again of sets of equations, we realize that f will often be quite complicated to evaluate, and that a good deal of computing time may thus be involved in each iteration.

A modification of Milne's method which counters this objection by requiring only two evaluations of f at each step can be used, at a modest cost in accuracy.

We change our notation slightly and write

$$p_n=\bar{y}_n;\ c_n=y_n. \tag{4.2.8}$$

Thus, we rewrite (4.2.1) as

$$p_{n+1}=y_{n-3}+\frac{4h}{3}(2y'_n-y'_{n-1}+2y'_{n-2}). \tag{4.2.9}$$

Also (4.2.7) becomes

$$c_{n+1}-p_{n+1}=\frac{29}{90}h^5y^v(\eta)\approx\frac{29}{28}T_1. \tag{4.2.10}$$

Let us assume that $y^v(\eta)$ is, to a fair approximation, constant from one step to the next. Then, using values from the n'th step, we would find, from (4.2.10),

$$T_1\approx\frac{28}{29}(c_n-p_n), \tag{4.2.11}$$

and so, adding T_1 to p_{n+1} should give a very good estimate of c_{n+1}. We write

$$m_{n+1}=p_{n+1}+\frac{28}{29}(c_n-p_n), \tag{4.2.12}$$

and

$$m'_{n+1} = f(x_{n+1}, m_{n+1}).$$

Now, using m'_{n+1}, we rewrite (4.2.3) as

$$c_{n+1} = y_{n-1} + \frac{h}{3}(m'_{n+1} + 4y'_n + y'_{n-1}). \qquad (4.2.13)$$

We can further improve this; again, from (4.2.7), we may write

$$T_2 \approx \frac{-1}{29}(c_{n+1} - p_{n+1}), \qquad (4.2.14)$$

and so, adding T_2 to c_{n+1}, we have, finally,

$$y_{n+1} = c_{n+1} + \frac{1}{29}(p_{n+1} - c_{n+1}), \qquad (4.2.15)$$

and

$$y'_{n+1} = f(x_{n+1}, y_{n+1}). \qquad (4.2.16)$$

The quantities p are called the "predictors", m the "modifiers" and c the "correctors."

This scheme involves the evaluation of f only twice at each step, and is, in practice, very nearly as accurate as the iterative scheme using (4.2.1), (4.2.2), and (4.2.3). Furthermore, we can use the quantity

$c_{n+1}-p_{n+1}$ as an estimate of the error at each step.

At this point, we return to the original problem of this section and describe this process algorithmically for the set of first order equations (3.3.4) subject to (3.3.6). To do so, we represent the vectors $\underline{y}_{n-3}$, $\underline{y}_{n-2}$, $\underline{y}_{n-1}$, and $\underline{y}_n$ as the columns $\underline{Y}_1$, $\underline{Y}_2$, $\underline{Y}_3$, and $\underline{Y}_4$ respectively of a matrix $\underline{Y}$. The vectors $\underline{y}'_{n-2}$, $\underline{y}'_{n-1}$, and $\underline{y}'_n$ we represent as the columns $\underline{W}_1$, $\underline{W}_2$, and $\underline{W}_3$ of a matrix $\underline{W}$. Also, the vectors $\underline{p}_n$ and $\underline{p}_{n+1}$ are represented as the columns $\underline{P}_1$ and $\underline{P}_2$ of the matrix $\underline{P}$. $\underline{\bar{\epsilon}}$ is a zero vector. It is assumed that some starting process has calculated and output the values of $\underline{Y}_1$, $\underline{Y}_2$, $\underline{Y}_3$, $\underline{Y}_4$, $\underline{W}_1$, $\underline{W}_2$, and $\underline{W}_3$, for $x = x_o$, x_o+h, x_o+2h, and x_o+3h respectively. The vectors $\underline{m}$,$\underline{n}$ and $\underline{c}$ correspond to the quantities m_{n+1}, m'_{n+1}, and c_{n+1} of equations (4.2.12), (4.2.12), and (4.2.13), respectively:

MILNE'S METHOD

1	$x \leftarrow 4h + x_o$	initialize x
2	$\underline{c} \leftarrow \underline{\bar{\epsilon}}$	initialize $\underline{c}$
3	$\underline{P}_1 \leftarrow \underline{\bar{\epsilon}}$	initialize $\underline{P}_1$
4	$\underline{P}_2 \leftarrow \underline{Y}_1+\frac{4h}{3}(2\underline{W}_3-\underline{W}_2+2\underline{W}_1)$	calculate predictors
5	$\underline{m} \leftarrow \underline{P}_2+\frac{28}{29}(\underline{c}-\underline{P}_1)$	calculate modifiers
6	$\underline{n} \leftarrow \underline{f}(x,\underline{m})$	calculate derivatives of modifiers

7	$\underline{c} \leftarrow \underline{Y}_3+\frac{h}{3}(\underline{n}+4\underline{W}_3+\underline{W}_2)$	calculate correctors
8	$\underline{Y}_1 \leftarrow \underline{Y}_2$	Shift columns of $\underline{Y}$ left one position.
9	$\underline{Y}_2 \leftarrow \underline{Y}_3$	
10	$\underline{Y}_3 \leftarrow \underline{Y}_4$	
11	$\underline{W}_1 \leftarrow \underline{W}_2$	Shift columns of $\underline{W}$ left one position.
12	$\underline{W}_2 \leftarrow \underline{W}_3$	
13	$\underline{Y}_4 \leftarrow \underline{c}+\frac{1}{29}(\underline{P}_2-\underline{c})$	calculate final values
14	$\underline{W}_3 \leftarrow \underline{f}(x,\underline{Y}_4)$	calculate final derivatives
15	$\underline{P}_1 \leftarrow \underline{P}_2$	shift columns of $\underline{P}$ left.
16	PRINT: x, $\underline{Y}_4$	output results
17	$x \leftarrow x+h$	increment h
18	$x:x_{max},(\leq,>)\rightarrow(4,19)$	test for final x
19	STOP.	

Inspection of this program shows that we are carrying four columns of $\underline{Y}$, three of $\underline{W}$, two of $\underline{P}$, and one each of $\underline{c}$, $\underline{m}$, and $\underline{n}$, for a total of 12ℓ computer storage locations, where ℓ is the number of simultaneous equations being solved. Acutally steps 5,6,7 could be written, somewhat awkwardly, as one, thus eliminating the vectors $\underline{m}$ and $\underline{n}$. Hence, we require 10ℓ storage locations.

It is possible, using $\underline{c}$ - $\underline{P}_1$ as a criterion to

gauge the truncation error at each step. Based on this, we can increase or decrease the interval h to gain speed when possible, or to preserve accuracy when necessary. Such a scheme is described in Ralston and Wilf [16], but requires storage for additional columns of $\underline{Y}$ and $\underline{W}$.

The routine described above should be fairly fast, the major time being spent in the two evaluations of $\underline{f}$.

Thus, Milne's method seems ideal, except for one thing. Nowhere in Milne's original paper [18], and, indeed, nowhere in his later book [20], is there any discussion at all of stability. In fact, the word does not appear to be used at all in either reference.

As we shall see in the following section, Milne's method suffers from definite instability, but, it can be modified at fairly small cost, to guarantee stability under fairly mild conditions.

Section 4.3 Stability of Multi-Step Methods

Recalling the example of Section 3.5, we saw that a component of an undesired solution of a difference equation could grow and dominate the desired solution. In that section, we were approximating a second-order

differential equation by a second-order difference equation. With the multi-step methods of the previous two sections, we have been approximating first order differential equations by higher order difference equations. Accordingly, we must be extremely wary lest one of the several solutions of such higher order difference equations misbehave and dominate the desired solution.

In this section, we present an analysis of the stability of Milne's method by a method which is of general applicability.

We will consider the iterative version of Milne's method, since the analysis is somewhat simpler than the non-iterative method, and the stability properties of the two are not greatly different.

Following Hamming [21], we consider the "generalized corrector" formula

$$y_{n+1} = py_n + qy_{n-1} + ry_{n-2}$$

$$+h(sy'_{n+1} + ty'_n + uy'_{n-1}), \qquad (4.3.1)$$

in place of (4.2.3).

Let z be the true solution of the given problem; that is

$$z' = \frac{dz}{dx} = f(x,z), \tag{4.3.2}$$

whereas the solution which we calculate satisfies

$$y'_n = f(x_n,y_n) + E(n), \tag{4.3.3}$$

where E(n) is the error at the n'th step, and is supposed to be small.

Thus, we can write

$$z_{n+1} = pz_n + qz_{n-1} + rz_{n-2}$$
$$+h(sz'_{n+1} + tz'_n + uz'_{n-1})+F(n), \tag{4.3.4}$$

where F(n) is the corresponding error.

We define ε, the error of the solution, as

$$\varepsilon_n = z_n - y_n. \tag{4.3.5}$$

Subtracting (4.3.1) from (4.3.4), we have

$$\varepsilon_{n+1} = p\varepsilon_n + q\varepsilon_{n-1} + r\varepsilon_{n-2}$$
$$+h(s\varepsilon'_{n+1} + t\varepsilon'_n + u\varepsilon'_{n-1}) + F(n) \tag{4.3.6}$$

Also, subtracting (4.3.3) from (4.3.2), we have

$$\varepsilon_n' = f(x_n,z_n)-f(x_n,y_n)-E(n). \qquad (4.3.7)$$

Now, applying the mean-value theorem to (4.3.7),

$$\varepsilon_n' = \frac{\partial f(x_n,\eta_n)}{\partial y}\,\varepsilon_n - E(n), \qquad (4.3.8)$$

for some $\eta_n \in [y_n,z_n]$.

In practice, E(n), F(n), and $\frac{\partial f}{\partial y}$ change fairly slowly from step to step, and so we will treat them as constants.

Writing

$$A = \frac{\partial f}{\partial y}\,, \qquad (4.3.9)$$

we transform the independent variable x, to

$$x = t/A. \qquad (4.3.10)$$

In these units, we may now rewrite (4.3.8) as

$$\frac{d\varepsilon}{dt} = \varepsilon - \frac{E}{A}, \qquad (4.3.11)$$

which we substitute into (4.3.6), to obtain

$$\varepsilon_{n+1} = p\varepsilon_n + q\varepsilon_{n-1} + r\varepsilon_{n-2}$$

$$+ k(s\varepsilon_{n+1} + t\varepsilon_n + u\varepsilon_{n-1})$$

$$+ F - \frac{k}{A}(s+t+u)E, \qquad (4.3.12)$$

where $k = hA$.

Equation (4.3.12) is a third-order linear difference equation, which we attempt to solve using the trial solution

$$\varepsilon_n = \rho^n, \qquad (4.3.13)$$

This yields the indicial equation

$$\rho^3 = p\rho^2+q\rho+r+k(s\rho^3+t\rho^2+u\rho), \qquad (4.3.14)$$

or

$$k = \frac{\rho^3-p\rho^2-q\rho-r}{s\rho^3+t\rho^2+u\rho} . \qquad (4.3.15)$$

For Milne's method, we evidently have

$$p=r=0,\ q=1,\ s=u=\tfrac{1}{3},\ t=\tfrac{4}{3}, \qquad (4.3.16)$$

whence (4.3.15) becomes

$$k = \frac{3(\rho^2-1)}{\rho^2+4\rho+1}, \qquad (4.3.17)$$

which, for any given k has two roots ρ_1 and ρ_2.

Thus, the general solution of (4.3.11) is

$$\varepsilon_n = c_1\rho_1^n + c_2\rho_2^n + \gamma, \qquad (4.3.16)$$

for some γ.

Hamming shows, by a graphical technique, that one or the other of ρ_1 of ρ_2 is always greater than 1 in absolute value whether A is positive or negative. Hence Milne's method is definitely unstable.

We point out also that, in practice, the absolute magnitude of the error is not the prime consideration; rather, we are usually more interested in the size of the error relative to the size of the solution. Thus, we are led to the concept of "relative stability," implying that the relative error of an approximation remains bounded. For relative stability, we may be more interested in knowing that some function such as ρe^{-h} is less than unity in absolute value.

In addition to demonstrating the instability of Milne's method, Hamming shows an acceptable alternative procedure. He observes that Milne's scheme is exact for a polynomial of degree four or less.

Hamming imposes the auxiliary conditions that the corrector formula (4.3.1) be exact for $y=1,x,x^2,x^3,x^4$. Thus, he obtains the relations

$$\left.\begin{aligned} p &= \frac{27(1-q)}{24} & s &= \frac{9-q}{24} \\ q &= q & t &= \frac{18+14q}{24} \\ r &= \frac{-3(1-q)}{24} & u &= \frac{-9+17q}{24} \end{aligned}\right\}. \qquad (4.3.17)$$

He further shows that the error term is

$$mh^5y^{V}(\eta) = \frac{-9+5q}{360}\, h^5y^{V}(\eta). \qquad (4.3.18)$$

For $h = k = 0$, (4.3.13) becomes

$$\rho^3-p\rho^2-q\rho-r = 0, \qquad (4.3.19)$$

or, using the values from (4.3.17),

$$8\rho^3-9(1-q)\rho^2-8q\rho+(1-q) = 0. \qquad (4.3.20)$$

Plotting (4.3.20) as a graph of ρ versus q, Hamming shows that the value q=0 (for Milne's method, q=1) provides a good compromise between the requirements of stability, and of ease of evaluation.

The method implied by q = 0 is often called "Hamming's Method." It is stable and relatively stable for A<0 if

$$h < \frac{.75}{|A|} , \qquad (4.3.20)$$

and is relatively stable for A>0 if

$$h < \frac{.4}{A} . \qquad (4.3.21)$$

A variation on Hamming's method, in the same vein as the version of Milne's method which does not require iteration, requires

$$h < \frac{.65}{|A|} .$$

This latter method may be represented algorithmically by replacing steps 5,7, and 13 in the algorithm for Milne's method by

---

5	$\underline{m} \leftarrow \underline{P}_2 + \frac{112}{121}(\underline{c} - \underline{P}_1)$
7	$\underline{c} \leftarrow \frac{1}{8}(9\underline{Y}_4 - \underline{Y}_2) + 3h(\underline{n} + 2\underline{W}_3 - \underline{W}_2)$
13	$\underline{Y}_4 \leftarrow \underline{c} + \frac{9}{121}(\underline{P}_2 - \underline{c})$

The procedure described above for analyzing the stability of the Milne process may be employed, with obvious modifications, to the analysis of many multi-step procedures. Detailed, and more general analyses are given by Urabe [22], and by Hull and Luxemburg [23].

Section 4.4 Runge-Kutta Methods for Initial Value Problems

The methods of the previous three sections are called multi-step mthods because they make use of information several steps back in the integration process. In this section, we consider several members of the class of one-step methods, so-called because they make no use of previous data.

We recall that one of the major objections to the Taylor series method was that expressions for derivatives of f are difficult to calculate. The Runge-Kutta methods avoid this difficulty by using additional evaluations of f itself, to attain comparable accuracy.

Following Henrici [9], we demonstrate the basic approach by developing a Runge-Kutta formula of the second order for the usual problem.

$$y'(x) = f(x,y), \qquad (4.1.1)$$

subject to

$$y(x_o) = y_o. \qquad (4.1.2)$$

By the Taylor series method, we can write

$$y(x+h)-y(x) = hy'(x) + \frac{h^2}{2!} y''(x) + O(h^3), \qquad (4.4.1)$$

or,

$$\frac{y(x+h)-y(x)}{h} = \Psi(x,y;h), \qquad (4.4.2)$$

where

$$\Psi(x,y;h) = f(x,y)+\frac{h}{2}f'(x,y)+O(h^2) \qquad (4.4.2).$$

By way of an approximation to $\Psi(x,y;h)$, let us consider

$$\Phi(x,y;h)= a_1 f(x,y)+a_2 f(x+p_1 h, y+p_2 hf(x,y)), \qquad (4.4.3)$$

with the parameters a_1, a_2, p_1, and p_2 at our disposal. Expanding in a Taylor series, we have

$$f(x+p_1h,y+p_2hf(x,y))$$

$$= f(x,y)+p_1h\frac{\partial f(x,y)}{\partial x}+p_2hf(x,y)\frac{\partial f(x,y)}{\partial y}+O(h^2), \qquad (4.4.4)$$

and thus we may rewrite Φ as

$$\Phi(x,y;h) = (a_1+a_2)f(x,y)+a_2h\left[p_1\frac{\partial f(x,y)}{\partial x}\right.$$

$$\left. + p_2 f(x,y)\frac{\partial f(x,y)}{\partial y}\right]+O(h^2). \qquad (4.4.5)$$

Also, by elementary calculus, and using (4.4.1),

$$f'(x,y) = \frac{d}{dx}f(x,y)$$

$$= \frac{\partial f(x,y)}{\partial x}+ f(x,y)\ \frac{\partial f(x,y)}{\partial y}, \qquad (4.4.6)$$

and thus we may rewrite Ψ as

$$\Psi(x,y;h) = f(x,y)+\frac{h}{2}\left[\frac{\partial f(x,y)}{\partial x}+f(x,y)\frac{\partial f(x,y)}{\partial y}\right]+O(h^2)$$

(4.4.7)

We now require that Φ and Ψ agree within $O(h^2)$. Thus, matching terms in (4.4.5) with (4.4.7), we have, from the constant terms,

$$a_1+a_2 = 1, \qquad (4.4.8)$$

and from the coefficient of h,

$$a_2p_1 = \frac{1}{2},\ a_2p_2 = \frac{1}{2}. \qquad (4.4.9)$$

Thus, writing $\alpha = a_2$ we have immediately,

$$a_1 = 1-\alpha,\ a_2 = \alpha,\ p_1 = p_2 = \frac{1}{2\alpha}\ , \qquad (4.4.10)$$

and α is free to take on any non-zero value. Hence, we write Φ as

$$\Phi(x,y;h) = (1-\alpha)f(x,y)+\alpha f\{x+\frac{h}{2\alpha},y+\frac{h}{2\alpha}f(x,y)\}, \qquad (4.4.11)$$

and note that Φ deviates from Ψ only by $O(h^2)$ for any value of α.

Two particular values of α are historically important:

(i) For $\alpha = 1/2$, we have the "improved Euler method" or the "Heun method." The approximation is written

$$y_{n+1} = y_n+\frac{h}{2}\{f(x_n,y_n)+f(x_n+h,y_n+hf(x_n,y_n))\}; \qquad (4.4.12)$$

(ii) For $\alpha = 1$, we have the "improved polygon method," or "modified Euler method," written

$$y_{n+1}=y_n+hf(x_n+\frac{h}{2},y_n+\frac{h}{2}f(x_n,y_n)). \qquad (4.4.13)$$

Notice once again that both of these formulas have the same order of truncation error as the Taylor series

method with one derivative term, but that they both achieve this by two evaluations of f, and no evaluations of the derivative of f.

The classical Runge-Kutta formula [24,25] is similar to the one we have discussed, but it employs a weighted sum of four evaluations of f, and thus has the same order of truncation error as a Taylor series with the third derivative term of f.

Specifically,

$$\Phi(x,y;h) = \mu_1 k_1 + \mu_2 k_2 + \mu_3 k_3 + \mu_4 k_4, \qquad (4.4.14)$$

where

$$\left.\begin{aligned} k_1 &= f(x,y) \\ k_2 &= f(x+\alpha h, y+\beta k_1) \\ k_3 &+ f(x+\alpha_1 h, y+\beta_1 k_1 + \gamma_1 k_2) \\ k_4 &= f(x+\alpha_2 h, y+\beta_2 k_1 + \gamma_2 k_2 + \delta_2 k_3) \end{aligned}\right\} \qquad (4.4.15)$$

Relations among the α's, β's, γ's, δ_2, and μ's are obtained by requiring the Taylor expansion of (4.4.14) to agree with the Taylor series method up to terms of

order h^4. The algebra involved in this process is quite lengthy, but involves only standard techniques, and is of no particular interest in itself; accordingly, we omit it from this discussion. The details may be found, for example, in Ralston and Wilf [16].

This process leads to eight equations in the thirteen parameters. By requiring these to be independent of f itself, one obtains

$$\left.\begin{aligned} \alpha &= \beta \\ \alpha_1 &= \beta_1+\gamma_1 \\ \alpha_2 &= \beta_2+\gamma_2+\delta_2 \end{aligned}\right\} \qquad (4.4.16)$$

and the following eight equations in ten parameters

$$\left.\begin{aligned} \mu_1+\mu_2+\mu_3+\mu_4 &= 1 \\ \mu_2\alpha+\mu_3\alpha_1+\mu_4\alpha_2 &= \frac{1}{2} \\ \mu_2\alpha^2+\mu_3\alpha_1^2+\mu_4\alpha_2^2 &= \frac{1}{3} \\ \mu_2\alpha^3+\mu_3\alpha_1^3+\mu_4\alpha_2^3 &= \frac{1}{4} \\ \mu_3\alpha\gamma_1+\mu_4(\alpha\gamma_2+\alpha_1\delta_2) &= \frac{1}{6} \\ \mu_3\alpha^2\gamma_1+\mu_4(\alpha^2\gamma_2+\alpha_1^2\delta_2) &= \frac{1}{12} \\ \mu_3\alpha\alpha_1\gamma_1+\mu_4(\alpha\gamma_2+\alpha_1\delta_2)\alpha_2 &= \frac{1}{8} \\ \mu_4\alpha\gamma_1\delta_2 &= \frac{1}{24} \end{aligned}\right\} \qquad (4.4.17)$$

Equations (4.4.17) have two degrees of freedom; most of the variations of the Runge-Kutta scheme make use of this freedom in slightly different ways.

Runge [24], used the following:

$$\left.\begin{aligned} \mu_1 &= \mu_4 = 1/6, \\ \mu_2 &= \mu_3 = 1/3, \\ k_1 &= f(x,y), \\ k_2 &= f(x+h/2, y+k_1/2), \\ k_3 &= f(x+h/2, y+k_2/2), \\ k_4 &= f(x+h, y+h). \end{aligned}\right\} \qquad (4.4.18)$$

Kutta [25], chose instead:

$$\left.\begin{aligned} \mu_1 &= \mu_4 = 1/8, \\ \mu_2 &= \mu_3 = 3/8, \\ k_1 &= f(x,y), \\ k_2 &= f(x+h/3, y+k_1/3), \\ k_3 &= f(x+\tfrac{2h}{3}, y-k_1/3+k_2), \\ k_4 &= f(x+h, y+k_1-k_2+k_3). \end{aligned}\right\} \qquad (4.4.19)$$

These choices appear to have been motivated mostly by a desire to make the algebra as simple as possible. More recently, Gill [26], motivated by the availability of a digital computer (EDSAC) with a small memory, and also by

a desire to minimize the round-off error, used the following, which is commonly known as the Runge-Kutta-Gill procedure, and is probably the most-used method for solving initial value problems using a computer:

$$\left.\begin{aligned}
k_1 &= f(x_o,y_o), & y_1 &= y_o+(k_1-2q_o)/2,\\
k_2 &= f(x_o+h/2,y_1), & y_2 &= y_1+(1-\sqrt{1/2})(k_2-q_1),\\
k_3 &= f(x_o+h/2,y_2), & y_3 &= y_2+(1+\sqrt{1/2})(k_3-q_2),\\
k_4 &= f(x_o+h,y_3); & y_4 &= y_3+(k_4-2q_3)/6;
\end{aligned}\right\} \quad (4.4.20)$$

$$\left.\begin{aligned}
q_1 &= q_o+\tfrac{3}{2}(k_1-2q_o)-k_1/2,\\
q_2 &= q_1+3(1-\sqrt{1/2})(k_2-q_1)-(1-\sqrt{1/2})k_2,\\
q_3 &= q_2+3(1+\sqrt{1/2})(k_3-q_2)-(1+\sqrt{1/2})k_3,\\
q_4 &= q_3+\tfrac{3}{6}(k_4-2q_3)-\tfrac{1}{2}k_4;
\end{aligned}\right\}$$

where $y(x_o+h) = y_4$, and $q_o = 0$ initially, and $q_o = q_4$ after the first step.

Clearly, as each of the k's, y's, and q's is calculated, it may replace its previous value.

We represent the Runge-Kutta-Gill process algorithmically for the set of equations

$$\underline{y}'(x) = \underline{f}(x,\underline{y}), \qquad (3.3.4)$$

subject to

$$\underline{y}(x_o) = \underline{\eta}. \qquad (3.3.6)$$

The vectors $\underline{y}$, $\underline{k}$ and $\underline{q}$ represent the obvious quantities of (4.4.20), and ℓ is the number of equations. We assume that the auxiliary vectors $\underline{a}$, $\underline{b}$, and $\underline{c}$ are initially

$$\underline{a} = (1/2,\ 1-\sqrt{1/2},\ 1+\sqrt{1/2},\ 1/6),$$
$$\underline{b} = (2,1,1,2),$$
$$\underline{c} = (1/2,\ 1-\sqrt{1/2},\ 1+\sqrt{1/2},\ 1/2).$$

RUNGE-KUTTA-GILL METHOD

→	1	$x \leftarrow x_o$	initialize x
	2	$\underline{y} \leftarrow \underline{\eta}$	initialize $\underline{y}$
	3	$\underline{q} \leftarrow \underline{\varepsilon}$	initialize $\underline{q}$
	4	$j \leftarrow 1$	initialize j
	5	$\underline{k} \leftarrow \underline{f}(x,\underline{y})$	evaluate $\underline{k}$
	6	$i \leftarrow 1$	initialize i
	7	$z \leftarrow a_j(k_i - b_j q_i)$	calculate auxiliary z
	8	$y_i \leftarrow y_i + hz$	calculate new y_i
	9	$q_i \leftarrow q_i + 3z - c_j k_i$	calculate new q_i
	10	$i \leftarrow i+1$	increment i
	11	$i:\ell,(\leq,>)\rightarrow(7,12)$	test
	12	$j \leftarrow j+1$	increment j

```
13   j:4,(≤,>)→(6,14)          test
14   PRINT;x,y                 output solution point
15   x←x+h                     increment x
16   x:x_max(≤,>)→(4,17)       test
17   STOP
```

We notice several things about the Runge-Kutta-Gill process, by way of comparison with Hamming's method. One great advantage of the Runge-Kutta scheme is that it is self-starting, unlike the Hamming procedure. On the other hand, it requires four evaluations of f for each step forward, and will therefore require about twice as long as Hammings method, on a computer. Although one of Gill's innovations was to reduce the storage required to 3ℓ locations (as opposed to 10ℓ for Hamming), this is no great advantage on modern computers which usually have large quantities of rapid, random-access memory.

Stability of Runge-Kutta methods is seldom discussed. However the method is sometimes unstable (see for example [11], page 91], although this can usually be overcome by taking h sufficiently small. A recent paper by Urabe [22] contains the most complete treatment known to us.

It is our experience that the Runge-Kutta method is over-used in many computing centres, on account of its ease of use. We would suggest that a much better general-purpose program would merely use a Runge-Kutta starter for a Hamming process, for the reasons which we have discussed above.

Section 4.5 Further Techniques for Boundary Value Problems

We again consider the linear, second order equation

$$y''(x)+f(x)y'(x)+g(x)y(x) = k(x), \qquad (3.6.3)$$

subject to

$$y(a) = \alpha, y(b) = \beta. \qquad (3.6.2)$$

Using the approximations

$$h^2D^2 = \delta^2\left(1- \frac{\delta^2}{12}+\frac{\delta^4}{90} - \frac{\delta^6}{360}+\ldots\right), \qquad (2.3.11)$$

and

$$hD = \mu\delta\left(1-\frac{\delta^2}{6}+ \frac{\delta^4}{30} - \ldots.\right), \qquad (2.3.15)$$

(3.6.3) becomes

$$\delta^2\left(1-\frac{\delta^2}{12}+\ldots\right)y_i+hf_i\mu\delta\left(1-\frac{\delta^2}{6}+\ldots\right)y_i+h^2g_iy_i=h^2k_i \ . \qquad (4.5.1)$$

That is

$$\delta^2 y_i + h f_i \mu\delta y_i + h^2 g_i y_i = h^2 k_i + C y_i, \qquad (4.5.2)$$

where

$$C = \left(\frac{\delta^4}{12} - \frac{\delta^6}{90} + \ldots\right) + h f_i \left(\frac{\mu\delta^3}{6} - \frac{\mu\delta^5}{30} + \ldots\right) \qquad (4.5.3)$$

Rewriting (4.5.2) in terms of ordinates,

$$y_{i-1}\left(1 - \frac{h}{2} f_i\right) - y_i\left(2 - h^2 g_i\right) + y_{i+1}\left(1 + \frac{h}{2} f_i\right) = h^2 k_i + C y_i. \qquad (4.5.4)$$

Thus, in the same manner as in Section 3.6, we can write (4.5.4) as

$$\underline{A}\ \underline{y} = \underline{d} - \underline{C}\ \underline{y}, \qquad (4.5.5)$$

and the boundary conditions are included in $\underline{d}$.

Now, for a first approximation, we neglect $\underline{C}\ \underline{y}$, and obtain a first approximation $\underline{y}^1$ to $\underline{y}$ by solving

$$\underline{A}\ \underline{y}^1 = \underline{d} \qquad (4.5.6)$$

We next difference $\underline{y}^1$ to obtain an initial estimate of $\underline{C}\ \underline{y}$, and calculate a correction, $\underline{\eta}^1$ to $\underline{y}^1$ by

$$\underline{A}\ \underline{\eta}^1 = -\underline{C}\ \underline{y}^1. \qquad (4.5.7)$$

If $\underline{C}\ \underline{\eta}^1$ is still large, we can calculate a further correction $\underline{\eta}^2$, from

$$\underline{A}\ \underline{\eta}^2 = -\underline{C}\ \underline{\eta}^1, \tag{4.5.7}$$

and so on until the corrections have become as small as we desire. Then

$$\underline{y} = \underline{y}^1+\underline{\eta}^1+\underline{\eta}^2+\ldots \quad . \tag{4.5.8}$$

The only difficulty with this process is that, in order to evaluate the difference corrections, using central differences, we require values of y outside the range [a,b]. These can readily be obtained by using equation (4.5.4) as a predictor formula.

An example of this process is given in [11], page 95.

Another way to solve the given boundary value problem is to replace it with a pair of initial value problems, as follows:

Using any of the methods which we have discussed for solving initial value problems, obtain a solution y_1, of (3.6.3) subject to

$$y_1(a) = \alpha,\ y_1'(a) = \gamma_1, \tag{4.5.9}$$

and obtain

$$y_1(b) = \beta_1. \tag{4.5.10}$$

Then, solve the homogeneous form of (3.6.3), that is

$$y_2''(x)+f(x)y_2'(x)+g(x)y_2(x) = 0, \tag{4.5.11}$$

subject to

$$y_2(a) = 0,\ y_2'(a) = \gamma_2. \tag{4.5.12}$$

Since the differential equation (3.6.3) is linear, any linear combination of y_1 and y_2 also satisfies it. In particular,

$$y = y_1(x)+ty_2(x), \tag{4.5.13}$$

not only satisfies (3.6.3), but also $y(a) = \alpha$, for any t.

Clearly,

$$\begin{aligned} y(b) &= y_1(b)+ty_2(b) \\ &= \beta_1+t\beta_2=\beta, \end{aligned} \tag{4.5.14}$$

and thus, solving (4.5.14) for t,

$$t = \frac{\beta-\beta_1}{\beta_2}. \tag{4.5.15}$$

Hence, the desired solution is

$$y(x) = y_1(x)+\frac{\beta-\beta_1}{\beta_2}\,y_2(x). \tag{4.5.16}$$

CHAPTER V

A NUMERICAL APPROACH TO THE GREEN'S FUNCTION

In this chapter we consider the matrix approximation to the boundary value problem given in Chapter III. We show that, as the size of the matrix increases, and the tabular value decreases, the approximation takes on the form of the solution given by the well-known Green's function method.

A proof of convergence of the approximation is given, and an analytic expression for the Green's function is obtained. This expression is believed to be new.

Section 5.1 Introduction

In this chapter, we will consider the linear boundary value problem

$$y''(x)+q(x)y(x) = k(x), \qquad (5.1.1)$$

subject to

$$y(0) = \alpha, y(1) = \beta. \qquad (5.1.2)$$

The differential equation (5.1.1) is not as restricted as it might at first appear, for we can usually write

$$y''(x)+p(x)y'(x)+q(x)y(x) = r(x), \qquad (5.1.3)$$

in the same form as (5.1.1), as we now show. In (5.1.3),

let

$$y(x) = u(x)v(x), \tag{5.1.4}$$

whence

$$y'(x) = u'(x)v(x)+u(x)v'(x), \tag{5.1.5}$$

and

$$y''(x) = u''(x)v(x)+2u'(x)v'(x)+v''(x). \tag{5.1.6}$$

Thus, (5.1.3) may be written

$$u''v+u'(2v'+pv)+u(pv'+qv) = r-v'', \tag{5.1.7}$$

and, to eliminate the first derivative term in u, we set

$$2v'+pv = 0, \tag{5.1.8}$$

and we can solve for v by separation of variables. We can write (5.1.8) as

$$\frac{dv}{dx} = -\frac{p}{2}v\,, \tag{5.1.9}$$

or

$$\frac{dv}{v} = -\frac{p}{2}\,dx, \tag{5.1.10}$$

and hence, integrating over a suitable range

$$\ln v = -\frac{1}{2}\int p\,dx, \tag{5.1.10}$$

or

$$v = \exp\{-\frac{1}{2}\int p\,dx\}. \tag{5.1.11}$$

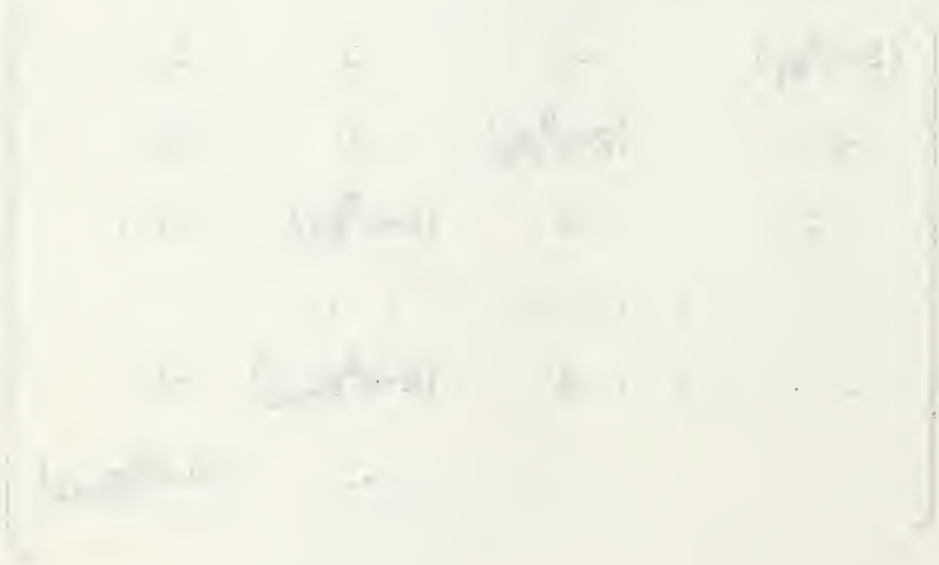
In this way, subject to fairly mild restrictions, equation (5.1.3) may be written in the form (5.1.1).

Let us now solve (5.1.1) numerically by the method of Section 3.6. Obviously, we again have to solve

$$\underline{A}\,\underline{y} = \underline{d}, \tag{3.6.14}$$

where now $\underline{A}$ is the matrix

$$-\begin{bmatrix} (2-h^2g_1) & -1 & 0 & 0\ldots \\ -1 & (2-h^2g_2) & -1 & 0\ldots \\ 0 & -1 & (2-h^2g_3) & -1\ldots \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & -1 & (2-h^2g_{n-2}) & -1 \\ & & -1 & (2-h^2g_{n-1}) \end{bmatrix}, \tag{5.1.12}$$



$\underline{y}$ is, as before

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_{n-1} \end{bmatrix}, \tag{5.1.13}$$

and

$$\underline{d} = \begin{bmatrix} h^2k_1-\alpha \\ h^2k_2 \\ h^2k_3 \\ . \\ . \\ . \\ h^2k_{n-2} \\ h^2k_{n-1}-\beta \end{bmatrix} = h^2 \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ . \\ . \\ . \\ k_{n-2} \\ k_{n-1} \end{bmatrix} - \begin{bmatrix} \alpha \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \\ \beta \end{bmatrix} = h^2\underline{k}-\underline{b}. \tag{5.1.14}$$

Thus, the solution is given by

$$\underline{y} = \underline{A}^{-1}(h^2\underline{k})-\underline{A}^{-1}\underline{b}. \tag{5.1.15}$$

Writing

$$\underline{G} = h\underline{A}^{-1}, \tag{5.1.16}$$

we have

$$\underline{y} = \underline{G}(h\underline{k}) - \underline{G}\,\underline{b}/h, \tag{5.1.17}$$

or, for the components of $\underline{y}$

$$y_i = \sum_{j=1}^{n-1} G_{ij}k_j h - \sum_{j=1}^{n-1} G_{ij}b_j/h. \tag{5.1.18}$$

But

$$b_j = \alpha\delta_{1,j} + \beta\delta_{n-1,j}, \tag{5.1.19}$$

where δ_{ij} is the Kroneker delta function, and so

$$y_i = \sum_{j=1}^{n-1} G_{ij}k_j h - \{\alpha\frac{G_{i,1}}{h} + \beta\frac{G_{i,n-1}}{h}\}. \tag{5.1.20}$$

Assuming for the moment that $\alpha = \beta = 0$, we have

$$y_i = \sum_{j=1}^{n-1} G_{ij}k_j h. \tag{5.1.21}$$

Now, it is well known [28] that the solution for equation (5.1.1) subject to (5.1.2) with $\alpha = \beta = 0$ can be written

$$y(x) = \int_0^1 G(x,\xi)k(\xi)d\xi, \tag{5.1.22}$$

where $g(x,\xi)$ is called the Green's function. We point out the obvious similarity between (5.1.21) and (5.1.22); although the one is a discrete formula, and the other continuous, it is clear that the sum term in the former corresponds to the integral term in the latter. In fact, (5.1.21) looks like a trapezoidal approximation to (5.1.22), and thus, we are led to suspect that, in the limit as $h \to 0$, (5.1.21) may coincide with (5.1.22); we will show that this is the case later.

We can go farther and use (5.1.20) to obtain a plausible extension to (5.1.22). We remove the requirement that $\alpha = \beta = 0$; now using the (so-far unproved) correspondence

$$G_{ij} \xrightarrow[h\to 0]{} G(x,\xi), \tag{5.1.23}$$

we obtain, by taking the limit of (5.1.21),

$$y(x) = \int_0^1 G(x,\xi)k(\xi)d\xi - \{\alpha \lim_{h\to 0} \frac{G(x,h)}{h} + \beta \lim_{h\to 0} \frac{G(x,1-h)}{h}\}. \tag{5.1.24}$$

Now, we know independently that $G(x,0)=G(x,1)=0$. Thus we may write the first term in the braces in (5.1.24)

as

$$\alpha \lim_{h\to 0} \frac{G(x,h)}{h} = \alpha \lim_{h\to 0} \frac{G(x,h)-G(x,0)}{h} \tag{5.1.25}$$

$$= \alpha \lim_{\xi\to 0} \frac{G(x,\xi)-G(x,0)}{\xi} = \alpha \left.\frac{\partial G(x,\xi)}{\partial \xi}\right|_{\xi=0}$$

$$= \alpha \frac{\partial G(x,0)}{\partial \xi}. \tag{5.1.26}$$

By a similar argument, the second term in the braces becomes

$$\beta \lim_{h\to 0} \frac{G(x,1-h)}{h} = -\beta \frac{\partial G(x,1)}{\partial \xi}. \tag{5.1.26}$$

Thus, we may write

$$y(x) = \beta \frac{\partial G(x,1)}{\partial \xi} - \alpha \frac{\partial G(x,0)}{\partial \xi} + \int_0^1 G(x,\xi)k(\xi)d\xi. \tag{5.1.27}$$

Formula (5.1.27) is a special case of Green's formula. In passing, we remark again that examination of numerical processes may often point toward a simple, intuitive approach to theorems of analysis (cf. our discussion of the fundamental theorem of Calculus, in Section 2.4).

The preceeding has been, at best, heuristic. We have provided no proof of the validity of our limiting argument, but we have nevertheless achieved some limited results. In the following section, we attempt to put these results on a sounder footing.

Section 5.2 A Proof of Convergence

In solving the problem (5.1.1), we replaced the differential equation by a matrix equation to obtain an approximate solution $\bar{\underline{y}}$, satisfying

$$\underline{A}\,\bar{\underline{y}} = \underline{d}. \tag{3.6.14}$$

The exact solution $\underline{y}$, satisfies, say

$$\underline{A}\,\underline{y} = \underline{d} + \underline{\varepsilon}. \tag{5.2.1}$$

We write

$$\underline{y} - \bar{\underline{y}} = \underline{\eta}, \tag{5.2.2}$$

and hence, substituting into (3.6.14), we obtain

$$\underline{A}\,\underline{\eta} = \underline{\varepsilon}. \tag{5.2.3}$$

Clearly, we would like to obtain a bound on $\underline{\eta}$ and show that $\underline{\eta} \to \underline{0}$ as $h \to 0$. To do so, we first need a

bound on $\underline{\varepsilon}$.

In deriving (3.6.14), we used the approximation

$$h^2 y_i'' \approx \delta^2 y_i = y_{i+1} - 2y_i + y_{i-1}. \tag{5.2.4}$$

Now, by Taylor's theorem,

$$y_{i+1} = y_i + hy_i' + \frac{h^2}{2!} y_i'' + \frac{h^3}{3!} y_i''' + \frac{h^4}{4!} y_i^{iv}(\xi_1), \tag{5.2.5}$$

$$-2y_i = -2y_i, \tag{5.2.6}$$

and,

$$y_{i-1} = y_i - hy_i' + \frac{h^2}{2!} y_i'' - \frac{h^3}{3!} y_i''' + \frac{h^4}{4!} y_i^{iv}(\xi_2). \tag{5.2.7}$$

Adding these three together, we have

$$\delta^2 y_i = \frac{2h^2}{2!} y_i'' + \frac{h^4}{4!} \{y_i^{iv}(\xi_1) + y_i^{iv}(\xi_2)\}. \tag{5.2.8}$$

Assuming that y is analytic, its derivatives are therefore bounded, and we may find some number γ satisfying

$$\frac{1}{4!} |y_i^{iv}(\xi_1) + y_i^{iv}(\xi_2)| < \gamma \tag{5.2.9}$$

uniformly for $x \in [0,1]$, or for $i \in [0,n]$. Thus, from (5.2.8),

$$\delta^2 y_i = h^2 y_i'' + \varepsilon_i \tag{5.2.10}$$

where, for each element of $\underline{\varepsilon}$, we have

$$|\varepsilon_i| < \gamma h^4. \tag{5.2.11}$$

Now, from (5.2.3),

$$\underline{\eta} = \underline{A}^{-1}\underline{\varepsilon}\ , \tag{5.2.12}$$

and therefore, for any element, η_i, of $\underline{\eta}$,

$$\eta_i = \sum_{\ell=1}^{n-1} A^{-1}_{i,\ell}\varepsilon_\ell. \tag{5.2.13}$$

Thus,

$$|\eta_i| \leq \sum_{\ell=1}^{n-1} |A^{-1}_{i,\ell}||\varepsilon_\ell|$$

$$< \gamma h^4 \sum_{\ell=1}^{n-1} |A^{-1}_{i,\ell}|. \tag{5.2.14}$$

Writing the expansion of $\underline{A}$ in terms of its eigenvalues and eigenvectors,

$$\underline{A} = \sum_{m=1}^{n-1} \lambda_m \underline{p}_m \underline{p}_m^t, \tag{5.2.15}$$

where $\underline{A}\,\underline{p}_m = \lambda_m \underline{p}_m$, and $\underline{p}_i^t \underline{p}_j = \delta_{i,j}$. (5.2.16)

Therefore,

$$\underline{A}^{-1} = \sum_{m=1}^{n-1} \frac{1}{\lambda_m} \underline{p}_m \underline{p}_m^t \tag{5.2.17}$$

and so

$$A^{-1}_{i,\ell} = \sum_{m=1}^{n-1} \frac{1}{\lambda_m} p_{m,i} p_{m,\ell}. \tag{5.2.18}$$

Accordingly, from (5.2.14),

$$|\eta_i| < \gamma h^4 \sum_{m=1}^{n-1} \left|\frac{1}{\lambda_m}\right| |p_{m,i}| \sum_{\ell=1}^{n-1} |p_{m,\ell}|, \tag{5.2.19}$$

but, it is well-known that

$$\sum_{\ell=1}^{n-1} |p_{m,\ell}| \leq \sqrt{n-1} < \sqrt{n} = 1/\sqrt{h} \tag{5.2.20}$$

Therefore,

$$|\eta_i| < \frac{\gamma h^4}{\sqrt{h}} \sum_{m=1}^{n-1} \left|\frac{1}{\lambda_m}\right| |p_{m,i}| . \qquad (5.2.21)$$

Suppose that the λ_i are numbered so that λ_1 is the smallest in absolute value; then

$$|\eta_i| < \frac{\gamma h^4}{\sqrt{h}} \frac{1}{|\lambda_1|} \sum_{m=1}^{n-1} |p_{m,i}|$$

$$< \frac{\gamma h^4}{\sqrt{h}} \cdot \frac{1}{|\lambda_1|} \frac{1}{\sqrt{h}} = \frac{\gamma h^3}{|\lambda_1|} . \qquad (5.2.22)$$

Thus, we see that unless $\lambda_1 \to 0$ faster than h^3, we will have $\underline{\eta} \to \underline{0}$, as $h \to 0$. It is not obvious whether this is restrictive or not; to get a clearer idea, we try an example. Let us suppose that in (5.1.1), $q(x) = 0$. Then, the matrix $\underline{A}$ becomes the well-known matrix $\underline{D}^2$, of which the smallest eigenvalue is [29]

$$\lambda_1 = 4 \sin^2\left(\frac{\Pi h}{2}\right) \qquad (5.2.23)$$

$$= 4\left\{\frac{\Pi h}{2} - \left(\frac{\Pi h}{2}\right)^3 + \left(\frac{\Pi h}{2}\right)^5 - \ldots\right\}^2 \qquad (5.2.24)$$

and for $h \to 0$,

$$\lambda_1 \to 4\frac{\Pi^2 h^2}{2^2} = \Pi^2 h^2. \tag{5.2.25}$$

Hence, for this case,

$$|\eta_i| < \frac{\gamma h^3}{\Pi^2 h^2} = \frac{\gamma h}{\Pi^2} \xrightarrow[h \to 0]{} 0. \tag{5.2.26}$$

This latter result encourages us to believe that, for most reasonable differential operators, convergence should be obtained with little difficulty.

Henrici [9] proves what amounts to a different, and somewhat weaker, formulation of this same result, in a form bearing more directly on the discussion of the preceeding section. We state his result, modified for our notation.

If $g(x)$ is twice continuously differentiable in $[0,1]$, and $g(x) \leq 0$, then there exists some constant K such that

$$|A^{-1}_{i,j} - G(ih, jh)| \leq h^2 K, \tag{5.2.27}$$

for h sufficiently small. Briefly, this implies that the correspondence which we drew between equations (5.1.21) and (5.1.22) is valid.

Section 5.3 An Analytic Expression for the Green's Function

In practice, as already indicated, the sytem (5.1.1) subject to (5.1.2) is solved by replacing it with

$$\underline{A}\ \underline{y} = \underline{d} = h^2\underline{k} - \underline{b} \qquad (3.6.14)$$

and using standard numerical techniques to solve these linear equations.

In this section, we will exploit the tri-diagonal feature of $\underline{A}$ to obtain an algebraic expression for $\underline{A}^{-1}$, and hence, to $\underline{G}$. Thence, we will obtain an expression for lim $h \to 0$ of $\underline{G}$.

Writing

$$\varepsilon_i = h^2 g_i, \qquad (5.3.1)$$

equation (5.1.12) becomes

$$-\underline{A} = \begin{bmatrix} (2-\varepsilon_1) & -1 & 0 & 0 & \cdots \\ -1 & (2-\varepsilon_2) & -1 & 0 & \cdots \\ 0 & -1 & (2-\varepsilon_3) & -1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & -1 & (2-\varepsilon_{m-1}) & -1 \\ & & & -1 & (2-\varepsilon_m) \end{bmatrix}, \qquad (5.3.2)$$

where

$$m = n-1. \tag{5.3.3}$$

We define $\Delta_o = 1$, and $\Delta_\ell(\alpha_1,\alpha_2,\ldots,\alpha_\ell)$ to be the determinant of an ℓ by ℓ matrix of the form of (5.3.2) with $(2-\varepsilon_{\alpha_1})$ in the first row, $(2-\varepsilon_{\alpha_2})$ in the second row, and so on. Then, we write the inverse of $-\underline{A}$ in terms of its co-factors. For m=2, we obviously have (since $\underline{G} = h\underline{A}^{-1}$

$$\underline{G} = \frac{-h}{\Delta_2(1,2)} \begin{bmatrix} \Delta_1(2) & \Delta_o \\ \Delta_o & \Delta_1(1) \end{bmatrix}; \tag{5.3.4}$$

for m = 3, we have

$$\underline{G} = \frac{-h}{\Delta_3(1,2,3)} \begin{bmatrix} \Delta_2(2,3) & \Delta_1(3) & \Delta_o \\ \Delta_1(3) & \Delta_2(1,3)+\Delta_o & \Delta_1(1) \\ \Delta_o & \Delta_1(1) & \Delta_2(1,2) \end{bmatrix}. \tag{5.3.5}$$

The elements of $\underline{G}$ increase rapidly in complexity as m increases, but by taking several more examples, a

pattern emerges. For m=4, we find

$$\underline{G}=\frac{-h}{\Delta_4(1,2,3,4)}\begin{vmatrix} \Delta_3(2,3,4) & \Delta_2(3,4) & \Delta_1(4) & \Delta_o \\ \Delta_2(3,4) & \Delta_3(1;3,4)+\Delta_1(4) & \Delta_2(1;4)+\Delta_o & \Delta_1(1) \\ \Delta_1(4) & \Delta_2(1;4)+\Delta_o & \Delta_3(1,2;4)+\Delta_1(1) & \Delta_2(1,2) \\ \Delta_o & \Delta_1(1) & \Delta_2(1,2) & \Delta_3(1,2,3) \end{vmatrix} \qquad (5.3.6)$$

It is clear from these examples that $G_{i,1}$ is given by

$$\frac{G_{i1}}{h} = \frac{-\Delta_{m-i}(i+1,i+2,\ldots,m)}{\Delta_m(1,2,\ldots,m)} \quad , \qquad (5.3.7)$$

and $G_{i,m}$ by

$$\frac{G_{i,m}}{h} = \frac{-\Delta_{i-1}(i-1,i,\ldots,m-1)}{\Delta_m(1,2,\ldots,m)} \quad , \qquad (5.3.8)$$

for any value of m. In a similar fashion, we find the general formula

$$\frac{\Delta_m(1,2,\ldots,m)}{h} \; G_{i,j} =$$

$$-\sum_{p=1}^{\min(m-i+1,\,j)} \Delta_{m-i+j-(2p-1)}(1,2,\ldots,j-p;i+p,i+p+1,\ldots,m), \tag{5.3.9}$$

for $j \leq i$,

and

$$G_{ij} = G_{ji}, \tag{5.3.10}$$

for $j > i$.

We now require an expression for $\Delta_\ell(\alpha_1,\alpha_2,\ldots,\alpha_\ell)$ in terms of the ε_i. Consider the expansion of Δ_ℓ in terms of its last row

$$\Delta_\ell(\alpha_1,\alpha_2,\ldots,\alpha_\ell) = (2-\varepsilon_{\alpha_\ell})\Delta_{\ell-1}(\alpha_1,\alpha_2,\ldots,\alpha_{\ell-1}) - \Delta_{\ell-2}(\alpha_1,\alpha_2,\ldots,\alpha_{\ell-2}). \tag{5.3.11}$$

Repeating this procedure, we attempt to write Δ_ℓ as a polynomial of the form

$$\Delta_\ell(\alpha_1,\alpha_2,\ldots,\alpha_\ell) = \ell+1 + \sum_{p=1}^{\ell} a_{\ell p}\,\varepsilon_{\alpha_p} + \sum_{p_2=1}^{p_2}\sum_{p_1=1}^{a_{\ell,p_1,p_2}}\varepsilon_{\alpha_{p_1}}\varepsilon_{\alpha_{p_2}}$$

$$+\sum_{p_3=1}^{\ell}\sum_{p_2=1}^{p_3}\sum_{p_1=1}^{p_2} a_{\ell,p_1,p_2,p_3}\,\varepsilon_{\alpha_{p_1}}\varepsilon_{\alpha_{p_2}}\varepsilon_{\alpha_{p_3}}+\ldots, \qquad (5.3.12)$$

where the a's are still to be determined. By comparing (5.3.12) with the expansion for the determinant Δ_ℓ, we find

$$\Delta_\ell(\alpha_1,\alpha_2,\ldots,\alpha_\ell) = \ell+1 - \sum_{p=1}^{\ell}(\ell+1-p)p\,\varepsilon_{\alpha_p}$$

$$+\sum_{p_2=1}^{\ell}\sum_{p_1=1}^{p_2}(\ell+1-p_2)(p_2-p_1)p_1\,\varepsilon_{\alpha_{p_1}}\varepsilon_{\alpha_{p_2}}$$

$$-\sum_{p_3=1}^{\ell}\sum_{p_2=1}^{p_3}\sum_{p_1=1}^{p_2}(\ell+1-p_3)(p_3-p_2)(p_2-p_1)p_1\varepsilon_{\alpha_{p_1}}\varepsilon_{\alpha_{p_2}}\varepsilon+\alpha_{p_3}\ldots \quad .$$

(5.3.13

We now - in view of later applications - consider the special case of ℓ=m of (5.3.13). We have

$$h\Delta_m(1,2,\ldots,m) = (m+1)h - \sum_{p=1}^{m}[(m+1)h-ph]phg(ph)h$$

$$+\sum_{p_2=1}^{m}\sum_{p_1=1}^{p_2}[(m+1)h-p_2h][p_2h-p_1h]p_1hg(p_1h)g(p_2h)hh$$

$$+\ldots \quad . \qquad (5.3.14)$$

Now, setting

$$u_i = p_1 h = p_i/(m+1) \tag{5.3.15}$$

we have

$$h\Delta_m(1,2,\ldots,m) = 1 - \sum_{u=h}^{1-h} [1-u]ug(u)h$$

$$+ \sum_{u_2=h}^{1-h} \sum_{u_1=h}^{u_2} [1-u_2][u_2-u_1]u_1 g(u_1)g(u_2)hh + \ldots \tag{5.3.16}$$

$$\xrightarrow[h \to 0]{} 1 - \int_{u=0}^{1} (1-u)ug(u)du$$

$$+ \int_{u_2=0}^{1} \int_{u_1=0}^{u_2} (1-u_2)(u_2-u_1)u_1 g(u_1)g(u_2)du_1 du_2 + \ldots \quad . \tag{5.3.17}$$

We must now consider another special case, the first column of the inverse matrix. We have, from (5.3.7)

$$\frac{\Delta_m}{h} G_{i1} = - \Delta_{m-i}(i+1,i+2,\ldots,m),$$

and hence, using (5.3.13)

$$- \Delta_m G_{11} = (m+1)h - ih - \sum_{p=1}^{m-i}{}' [(m+1)h - ih - ph]phg(ph+ih)h$$

$$+ \sum_{p_2=1}^{m} \sum_{p_1=1}^{p_2} [(m+1)h - ih - p_2h][p_2h - p_1h]p_1hg(p_1h+ih)$$

$$g(p_2h+ih)hh - \ldots, \qquad (5.3.18)$$

$$\xrightarrow[h \to 0]{} 1-x - \int_{u=0}^{1-x} (1-u-x)ug(u+x)du$$

$$+ \int_{u_2=0}^{1-x} \int_{u_1=0}^{u_2} (1-u_2-x)(u_2-u_1)u_1 g(u_1+x)g(u_2+x)du_1 du_2$$

$$- \ldots, \qquad (5.3.19)$$

which we write as $\Phi(x)$.

Clearly then, from (5.3.17),

$$h\Delta_m(1,2,\ldots,m) \to \Phi(0), \qquad (5.3.20)$$

and so we write

$$- \frac{G_{i,1}}{h} = \frac{h\Delta_{m-i}}{h\Delta_m} \to \frac{\Phi(x)}{\Phi(0)} = \Omega(x). \qquad (5.3.21)$$

Evidently,

$$\Omega(0) = \frac{\Phi(0)}{\Phi(0)} = 1, \tag{5.3.22}$$

and also, from (5.3.19), since, for x = 1, the integrals are over a range of zero length,

$$\Omega(1) = 0. \tag{5.3.23}$$

It is fairly clear that, by a process similar to that just described, we will also have

$$-\frac{G_{i,m}}{h} \rightarrow \Omega(1-x). \tag{5.3.24}$$

From (5.1.24) and (5.1.27), we see that $\Omega(x)$ from (5.3.21) is

$$\Omega(x) = \frac{\partial G(x,0)}{\partial \xi}, \tag{5.3.25}$$

and that $\Omega(1-x)$, from (5.3.24) is

$$\Omega(1-x) = -\frac{\partial G(x,1)}{\partial \xi} \tag{5.3.26}$$

Therefore, we see that

$$\omega_1(x) = \Omega(x), \tag{5.3.27}$$

and

$$\omega_2(x) = \Omega(1-x), \tag{5.3.28}$$

are two independent solutions of

$$\omega'' + g\omega = 0 \tag{5.3.29}$$

subject to

$$\omega_1(0) = \omega_2(1) = 1 \tag{5.3.30}$$

and

$$\omega_1(1) = \omega_2(0) = 0. \tag{5.3.31}$$

Hence, according to a standard result [28] we can now write

$$G(x,\xi) = \begin{cases} \frac{-1}{W}\,\omega_2(x)\omega_1(\xi) \ , & x \leq \xi \\ \frac{-1}{W}\,\omega_2(\xi)\omega_1(x) \ , & x \geq \xi, \end{cases} \tag{5.3.32}$$

where W is Wronskian

$$W = \omega_1\omega_2' - \omega_2\omega_1' \ , \tag{5.3.33}$$

and is a constant. We evaluate W at x = 1, and hence

$$W = - \omega_1'(1) = - \Omega'(1) = \frac{-\Phi'(1)}{\Phi(0)} , \qquad (5.3.34)$$

but $\Phi'(1)$ is easily seen to be -1, and so

$$W = = 1/\Phi(0). \qquad (5.3.35)$$

We thus rewrite (5.3.32)

$$G(x,\xi) = \begin{cases} \dfrac{\Phi(1-x)\Phi(\xi)}{\Phi(0)} , & x \leq \xi \\ \dfrac{\Phi(x)\Phi(1-\xi)}{\Phi(0)} , & x \geq \xi, \end{cases} \qquad (5.3.36)$$

where $\Phi(x)$ is given by (5.3.19).

The line of argument throughout this section has not been rigorous; in particular, we have provided no guarantee that the limits we have written actually exist. We gain some confidence (but not proof) from the theorem of the preceeding section; we will attempt further to make the argument seem plausible by means of the examples of the next section.

Section 5.4 Special Cases of the General Formula

We shall verify only that the ω's defined by (5.3.27) and (5.3.28) actually satisfy (5.3.29) to (5.3.31)

in some particular instances, since the work beyond that point is all well-known. It would be more satisfying to complete the anlytical investigation of the difficult limit processes involved, but this does not appear tractable at this stage.

First, we try $g(x)\equiv 0$. Then, we have

$$\omega_1(x) = \Omega(x) = \Phi(x)/\Phi(0), \tag{5.4.1}$$

and evaluating (5.3.19) with $g = 0$, all the integral terms vanish and we are left with

$$\Phi(x) = 1-x, \tag{5.4.2}$$

and hence

$$\Phi(0) = 1. \tag{5.4.3}$$

We test $\omega_1 = 1-x$ in the differential equations:

$$\omega_1' = -1, \tag{5.4.4}$$

and

$$\omega_1'' = 0, \tag{5.4.5}$$

and so (5.3.29) is satisfied. Furthermore,

$$\omega_1(0) = 1, \text{ and } \omega_1(1) = 0, \tag{5.4.6}$$

and so (5.3.30) and (5.3.31) are satisfied.

Also,

$$\omega_2(x) = \Omega(1-x) = \Phi(1-x)/\Phi(0)$$

$$= 1-(1-x) = x \tag{5.4.7}$$

and, as before, the required conditions are satisfied.

We now try the slightly more difficult function, $g(x) \equiv 1$. Thus, (5.3.29) becomes

$$\omega''+\omega = 0. \tag{5.4.8}$$

Now, from (5.3.19), with $g \equiv 1$, we have

$$\Phi(x) = 1-x - \int_{u=0}^{1-x}(1-u-x)u\ du$$

$$+ \int_{u_2=0}^{1-x}\int_{u_1=0}^{u_2}(1-u_2-x)(u_2-u_1)u_1 du_1 du_2 - \ldots \quad . \tag{5.4.9}$$

The first integral term is

$$I_1 = \int_{u=0}^{1-x}(1-u-x)u\ du = (1-x)\int_{u=0}^{1-x}u du - \int_{u=0}^{1-x}u^2 du$$

$$= (1-x)\left[\frac{u^2}{2}\right]_0^{1-x} - \left[\frac{u^3}{3}\right]_0^{1-x}$$

$$= \frac{(1-x)^3}{2} - \frac{(1-x)^3}{3} = \frac{(1-x)^3}{6} \quad . \qquad (5.4.10)$$

The second integral term is

$$I_2 = \int_{u_2=0}^{1-x}(1-u_2-x)\int_{u_2 0}^{u_2}(u_2-u_1)u_1 du_1 du_2$$

$$= \int_{u_2 0}^{1-x}(1-u_2-x)\frac{u_2^3}{6}du_2 = (1-x)\int_{u_2=0}^{1-x}\frac{u_2^3}{6}du_2 - \int_{u_2=0}^{1-x}u_2^4 du_2$$

$$= \frac{(1-x)^5}{24} - \frac{(1-x)^5}{30} = \frac{(1-x)^5}{120} = \frac{(1-x)^5}{5!} \quad . \qquad (5.4.11)$$

It can be shown that

$$\Phi(x) = (1-x) - \frac{(1-x)^3}{3!} + \frac{(1-x)^5}{5!} - \ldots, \qquad (5.4.12)$$

which we recognize as the Taylor series for $\sin(1-x)$; hence

$$\left.\begin{aligned} &\Phi(x) = \sin(1-x); \\ &\Phi(0) = \sin(1); \\ &\Phi(1-x) = \sin(x). \end{aligned}\right\} \qquad (5.4.13)$$

Thus,

$$\omega_1(x) = \sin(1-x)/\sin(1), \qquad (5.4.14)$$

and

$$\omega_2(x) = \sin(x)/\sin(1). \qquad (5.4.15)$$

Now,

$$\omega_1'(x) = -\cos(1-x)/\sin(1), \qquad (5.4.16)$$

and,

$$\omega_1''(x) = -\sin(1-x)/\sin(1), \qquad (5.4.17)$$

and so,

$$\omega_1'' + \omega_1 = 0 \qquad (5.4.18)$$

and (5.3.29) is satisfied. Clearly

$$\omega_1(0) = \sin(1)/\sin(1) = 1, \qquad (5.4.19)$$

and

$$\omega_1(1) = \sin(0)/\sin(1) = 0, \qquad (5.4.20)$$

satisfying (5.3.30) and (5.3.31).

The details for $\omega_2(x)$ are equally trivial. We have thus obtained concrete verification of the formulas of the previous section, and can feel fairly confident

that the analysis performed there is correct, and will probably work for more complicated functions, $g(x)$, although the integrations may become very difficult to perform. It may be that a possible approach is to approximate $g(x)$ by a polynomial, and then appeal to the Weierstrass approximation theorem, but even polynomials render the integrations in (5.3.19) very difficult.

It is unlikely that the formulas which we have found are in any way practical devices for finding analytic solutions to equations. They may, however, have some value in finding numerical solutions.

CHAPTER VI

CONCLUDING REMARKS

In the preceeding chapters, we have reviewed a narrow, but important, spectrum of numerical techniques for solving differential equations, and have discussed in some detail certain difficulties - such as instability - which hamper us in our attempts to obtain greater precision in the numerical results. Partial differential equations have been treated very briefly, but at sufficient length to show that each of the three main types requires a different numerical approach - one which is suggested by the analytic character of the differential system itself. In one chapter, we have examined the analytical behaviour of the finite-difference analogue of an ordinary differential equation subject to boundary conditions, as the tabular interval, h, approaches zero. The analysis is incomplete at present, and may be regarded as being of mostly heuristic value.

The numerical study of differential equations may seem to the mathematician nothing but a hodge-podge of special devices. It may, however, be said with some confidence that progress has now brought us to a point at which very few ordinary differential equations present any

serious difficulties to the numerical analyst. Probably, little more than this should be expected at this time, because the intensive study of the numerical solution of ordinary differential equations began only recently, and because it has usually been driven by the day-to-day necessity of solving this or that particular problem. It need hardly be emphasized that this situation is unsatisfactory.

Von Neumann once remarked that large digital computers could be justified simply on the basis that numerical studies of non-linear differential systems might lend some insight into those topics which ought to be studied by an analyst; or, as Hamming has said more recently, "The purpose of computing is insight, not numbers." Von Neumann's hope has not yet been fulfilled, but his remark does explain, and to a large extent justify, the energy expended by numerical workers on particular problems.

BIBLIOGRAPHY

[1] W.G. Bickley, "Difference and Associated Operators with some Applications," J. Math. and Phys., v. 27, Oct. 1948, p. 183.

[2] "Interpolation and Allied Tables," H.M. Stationery Office, 1956.

[3] Peirce and Foster, "A Short Table of Integrals," Ginn, 4 th. ed., 1956.

[4] The Dartmouth College Writing Group, "Modern Mathematical Methods and Models," v.1, Cushing-Malloy, 1958.

[5] R.W. Hamming, "Numerical Methods for Scientists and Engineers," McGraw, 1962.

[6] Johnson and Kiokemeister, "Calculus with Analytic Geometry," Allyn and Bacon, 1957.

[7] L. Euler, "Opera Omnia, Series Prima," v. 11 and 12, Leipzig and Berlin, 1913 and 1914.

[8] A.L. Cauchy, " Sur l' intégration des Équations Différentielles," lithographed at Prague, 1835.

[9] P. Henrici, "Discrete Variable Methods in Ordinary Differential Equations," Wiley, 1962.

[10] F.G. Tricomi, "Differential Equations," Hafner, 1961.

[11] "Notes on Applied Science No. 16: Modern Computing Methods," H.M. Stationery Office, 2nd ed., 1961.

[12] I.G. Petrovsky, "Lectures on Partial Differential Equations," Interscience, 1961.

[13] Courant and Hilbert, "Methods of Mathematical Physics," v. II, Interscience, 1962.

[14] R.S. Varga, "Matrix Iterative Analysis," Prentice-Hall, 1962.

[15] W.A. Murrary, "The Numerical Solution of Partial Differential Equations by Closed Difference Methods," master's thesis, University of Alberta, Oct., 1962.

[16] Ralston and Wilf, "Mathematical Methods for Digital Computers," Wiley, 1960.

[17] Birkhoff and Rota, "Ordinary Differential Equations," Ginn, 1962.

[18] W.E. Milne, "Numerical Integration of Ordinary Differential Equations," Am. Math. Monthly, v. 33, 1926, p. 455.

[19] Forsythe and Wasow, "Finite-difference Methods for Partial Differential Equations," Wiley, 1960.

[20] W.E. Milne, "Numerical Solution of Differential Equations," Wiley, 1953.

[21] R.W. Hamming, "Stable Predictor-corrector Methods," J. ACM, v. 6, 1959, p. 37.

[22] M. Urabe, "Theory of Errors in Numerical Integration of Ordinary Differential Equations," J. of Science of the Hiroshima University, v. 25, 1961, p. 3.

[23] T. Hull and W. Luxemberg, "Numerical Methods and Existence Theorems for Ordinary Differential Equations, "Num. Math., v. 2, 1960, p. 30.

[24] C. Runge, "Über die Numerische Auflösung von Differentialgleichungen, "Math. Ann., v. 46, 1895, p. 167.

[25] W. Kutte, "Beitrag zur näherungsweisen Integration totaler Differentialgleichungen, "Z.Math. Phys. v. 46, 1901, p. 435.

[26] S. Gill, "A process for the Step-by-step Integration of Differential Equations in an Automatic Digital Computing Machine, "Proc. Camb. Phil. Soc., v. 47, 1951, p. 96.

[27] K. Iverson, "A Programming Language, "Wiley, 1962.

[28] F. Hildebrand, "Methods of Applied Mathematics," Prentice-Hall, 1952.

[29] W.G. Bickley and J. McNamee, "Eigenvalues and Eigenfunctions of Finite-difference Operators, "Proc. Camb. Phil. Soc., v. 57, 1961, p. 532.

B29812